

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 14:39:32 ; Search time 539.84 Seconds

(without alignments)
135.498 Million cell updates/sec

Title: US-09-325-095-12

Perfect score: 23

Sequence: 1 AARATGATGATHTTTCATAC 23

Scoring table: IDENTITY_NUC

Searched: 679419 segs, 1590154680 residues

Database: GenEmbl.*

- 1: gb_pal.*
- 2: gb_ba2.*
- 3: gb_com.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pl3.*
- 10: gb_pl4.*
- 11: gb_pl5.*
- 12: gb_pl6.*
- 13: gb_pl7.*
- 14: gb_pl8.*
- 15: gb_pl9.*
- 16: gb_pl10.*
- 17: gb_pl11.*
- 18: gb_pl12.*
- 19: gb_pl13.*
- 20: gb_pl14.*
- 21: gb_pl15.*
- 22: gb_pl16.*
- 23: gb_pl17.*
- 24: gb_pl18.*
- 25: gb_pl19.*
- 26: gb_pl20.*
- 27: gb_pl21.*
- 28: gb_pl22.*
- 29: gb_pl23.*
- 30: gb_pl24.*
- 31: gb_pl25.*
- 32: gb_pl26.*
- 33: gb_pl27.*
- 34: gb_pl28.*
- 35: gb_pl29.*
- 36: gb_pl30.*
- 37: gb_pl31.*
- 38: gb_pl32.*
- 39: gb_pl33.*
- 40: gb_pl34.*
- 41: gb_pl35.*
- 42: gb_pl36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	20.6	89.6	3207 3 BOVPHOS3KN	M93252 Bovine phop

2	20.6	89.6	3452	4	AF001076
3	20.6	89.6	3498	5	A37232
4	20.6	89.6	23	5	A37240
5	20.6	89.6	3424	10	HSPH13K
6	20.6	89.6	3207	10	HSU79143
7	20.6	89.6	3207	12	MMU03279
8	20.6	89.6	3389	17	AF001075
9	18.6	80.9	5292	1	SEU77778
10	18	78.3	310617	34	HS69B13
11	18	78.3	5392	36	DMU12634
12	18	78.3	5381	37	DMU34777
13	17.8	77.4	199046	35	AC007744
14	17.4	75.7	108878	11	AC003092
15	17.4	75.7	259474	11	HUAC004605
16	17	73.9	4373	1	SGU61158
17	17	73.9	1777	7	LEPLEC14B
18	17	73.9	209618	10	HS54B20
19	17	73.9	169765	11	AC006075
20	17	73.9	2889	12	RATBEM1
21	17	73.9	4871	12	RNU28938
22	17	73.9	4524	36	CELC46E2
23	17	73.9	30250	36	CET26G10
24	17	73.9	124170	42	AC006000
25	17	73.9	193772	42	AC007073
26	16.8	73.0	82347	7	AB028614
27	16.8	73.0	140679	9	HS246H3
28	16.8	73.0	59765	11	AC005179
29	16.8	73.0	1000	14	HS246H3
30	16.8	73.0	186086	35	AC007165
31	16.8	73.0	184269	35	AC007271
32	16.8	73.0	81053	37	AC005762
33	16.6	72.2	251124	11	HUAE000660
34	16.4	71.3	1359	1	CKY16137
35	16.4	71.3	2370	1	CTENR
36	16.4	71.3	40245	1	MLCB8533
37	16.4	71.3	27322	1	MTCV261
38	16.4	71.3	42157	1	UO0017
39	16.4	71.3	6429	2	AF088800
40	16.4	71.3	40892	5	A59672
41	16.4	71.3	94301	7	ATF28A21
42	16.4	71.3	7074	7	SCYUR152W
43	16.4	71.3	2354	7	YSCDNL5A
44	16.4	71.3	86004	9	HS51723
45	16.4	71.3	103933	10	HS684024

ALIGNMENTS

RESULT 1	BOVPHOS3KN	3207 bp	MAW
LOCUS	BOVINE PHOSPHATIDYLINOSITOL 3-KINASE 110 KDA SUBUNIT MRNA, complete cds.		
DEFINITION	BOVINE PHOSPHATIDYLINOSITOL 3-KINASE 110 KDA SUBUNIT MRNA, complete cds.		
ACCESSION	M93252		
NID	G163519		
VERSION	M93252.1 GI:163519		
KEYWORDS	Phosphatidylinositol 3-kinase.		
SOURCE	Bos taurus CDNA to mRNA.		
ORGANISM	Bos taurus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bos.		
AUTHORS	Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R., Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtneidge, S.A., Parker, P.J., and Waterfield, M.D.		
TITLE	Phosphatidylinositol 3-kinase: Structure and expression of the 110 kd catalytic subunit		
JOURNAL	Cell 70, 419-429 (1992)		
MEDLINE	92354059		
FEATURES	Location/Qualifiers		
SOURCE	1..3207		

CDS

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/ec_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AAB30698.1"
/db_xref="GI:163520"
/db_xref="GI:163520"
/translation="MPPRPSGSELMGHIHMPRIIVTECLPENGMIYVTECLREATLT
IKHELFKARKYPIYLIQIDSSYIFSVYTOEAREEPEFDERRLDCLDLPDPFKVI
EVEGNEEKILNREIGFALGMPCEDMDKDEPNDFRNILNCKEADNLIDNSPH
SRAMYPPNVSSEPELPHKHTIKNDKQIIVIVIVSPNNDKQYTKIKHNDVPE
QVIAEIRKRTSMLSLSSEQLKCYLEQKYLKVCQDEYLFLEKYPISQYKIRSC
IMLGRPNMLMAKESLSQLPDCTMPSYSRISTATPYNGESTLSLWINSAL
RIKILCAIVVNIIDIKIVYRTGIYHGEPLCDNVTPRPGSNPRNEMLVNDIY
IPDLPRARILCLISICVGRKGAKEHCPLAMGNINLFDYDILVSGKALNLMVPH
GLEDLNPITGVTSNPNKEPCLTELEFDPFSSVVRPKMSVIEEHANMVSREAGSY
SHAGLSNRLARDNELRENDKEQRLACTDPSSETEQKEDFLSHRHVCVTIPELP
KILSVKNSRDEVAQMYCLVDPPIKEQAMELDCNYPDPVAGFAVRLKTYLT
DKLSQYLIQIVQVLYKEQYLDNOLVRELKALNORIGHEFFMHLKSEMNKTVSQ
RGLLESYCRACGMYLKLNRQVEAMEKILNLDILKOEKDETOKVOMKFLVQMR
RPFMDALQGFSPLPNPAHQNLRLSECRINMSAKRPLMNMENDIMSELFOENE
IIFKNGDDLRODMLTLQITRIMENIMONGIDLRLMYPGCLSIGCVGLIEVVRNHT
IMQIOCKGKLKALQNSHTLHOMLKDKRKGEIYDAIDILFTRSCAGVAFPIIGIG
DHNNSINWKGQGFHIDFGHLDHKKRFGKREVRPFVLQDFLIVISGADECT
KTRERFQEMCYKAYLAIROHANLEINLFSMILSGMPELOSFDIAYIRKTLALDK
TEOEALFYMKOMDANHGGWTKMDWIFHTIKOHALN"
BASE COUNT      1028 a      581 c      680 g      918 t
ORIGIN

```

```

Query Match      89.6%; Score 20.6; DB 3; Length 3207;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AARATGAYTGATHTTTCAYAC 23
11:|||||:|||||:|||||:|||||:
Db      3160 AAAATGATTGATCTTCACAC 3182

```

```

RESULT      2
LOCUS      AF001076      3452 bp      mRNA      VRT      08-JUL-1997
DEFINITION Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
complete cds.
ACCESSION  AF001076
NID        92245505
VERSION    AF001076.1 GI:2245505
KEYWORDS
SOURCE     Chicken.
ORGANISM   Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE  1 (bases 1 to 3452)
AUTHORS   Chang,H.W., Aoki,M., Fruman,D., Anger,K.R., Bellacosa,A.,
Tsichlis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
TITLE     Transformation of chicken cells by the gene encoding the catalytic
subunit of PI 3-kinase
JOURNAL    Science 276 (5320), 1848-1850 (1997)
MEDLINE    97334438
AUTHORS   Chang,H.W., Aoki,M. and Vogt,P.K.
TITLE     Direct Substitution
JOURNAL    Submitted (23-APR-1997) Molecular & Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
BCC-239, La Jolla, CA 92037, USA
FEATURES   source
1..3452
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="brain"

```

5'UTR
gene

CDS

```

1..150
151..3354
/gene="c-p3k"
/note="p3k proto-oncogene"
151..3357
/note="c-p3k"
/codon_start=1
/product="phosphoinositide 3-kinase catalytic subunit"
/protein_id="AAB62534.1"
/db_xref="PID:92245506"
/db_xref="GI:2245506"
/translation="MPPRPSGSELMGHIHMPRIIVTECLPENGMIYVTECLREATLT
IKHELFKARKYPIYLIQIDSSYIFSVYTOEAREEPEFDERRLDCLDLPDPFKVI
EVEGNEEKILNREIGFALGMPCEDMDKDEPNDFRNILNCKEADNLIDNSPH
SRAMYPPNVSSEPELPHKHTIKNDKQIIVIVIVSPNNDKQYTKIKHNDVPE
QVIAEIRKRTSMLSLSSEQLKCYLEQKYLKVCQDEYLFLEKYPISQYKIRSC
IMLGRPNMLMAKESLSQLPDCTMPSYSRISTATPYNGESTLSLWINSAL
RIKILCAIVVNIIDIKIVYRTGIYHGEPLCDNVTPRPGSNPRNEMLVNDIY
IPDLPRARILCLISICVGRKGAKEHCPLAMGNINLFDYDILVSGKALNLMVPH
GLEDLNPITGVTSNPNKEPCLTELEFDPFSSVVRPKMSVIEEHANMVSREAGSY
SHAGLSNRLARDNELRENDKEQRLACTDPSSETEQKEDFLSHRHVCVTIPELP
KILSVKNSRDEVAQMYCLVDPPIKEQAMELDCNYPDPVAGFAVRLKTYLT
DKLSQYLIQIVQVLYKEQYLDNOLVRELKALNORIGHEFFMHLKSEMNKTVSQ
RGLLESYCRACGMYLKLNRQVEAMEKILNLDILKOEKDETOKVOMKFLVQMR
RPFMDALQGFSPLPNPAHQNLRLSECRINMSAKRPLMNMENDIMSELFOENE
IIFKNGDDLRODMLTLQITRIMENIMONGIDLRLMYPGCLSIGCVGLIEVVRNHT
IMQIOCKGKLKALQNSHTLHOMLKDKRKGEIYDAIDILFTRSCAGVAFPIIGIG
DHNNSINWKGQGFHIDFGHLDHKKRFGKREVRPFVLQDFLIVISGADECT
KTRERFQEMCYKAYLAIROHANLEINLFSMILSGMPELOSFDIAYIRKTLALDK
TEOEALFYMKOMDANHGGWTKMDWIFHTIKOHALN"
BASE COUNT      1073 a      672 c      783 g      924 t
ORIGIN

```

```

Query Match      89.6%; Score 20.6; DB 4; Length 3452;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 AARATGAYTGATHTTTCAYAC 23
11:|||||:|||||:|||||:|||||:
Db      3310 AAAATGACTGATCTTCACAC 3332

```

```

RESULT      3
LOCUS      A37232      3498 bp      DNA      PAT      05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION  A37232
NID        92284345
VERSION    A37232.1 GI:2294345
KEYWORDS
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 3498)
AUTHORS   Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE     EUCHARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL    Patent: WO 9403609-A 1 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT    Other publication JP 8503124T 960409.
FEATURES   source
1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="PID:92294346"
/db_xref="GI:2294346"

```

Query Match	Best Local Similarity	Matches	18: Conservative	5: Mismatches	0: Indels	0: Gaps	0: Gaps
QY 1 AARATGAGTGGATHTTTCAYAC 23	89.6% Score 20.6; DB 5; Length 3498;	1147 a	626 c	722 g	- 1003 t		
Db 3160 AAAATGATGTGATCTTCACAC 3182							
RESULT 4	A37240	A37240	23 bp	DNA	PAT	05-MAR-1997	
LOCUS	DEFINITION	Sequence 9 from Patent WO9403609.					
ACCESSION	A37240						
NID	92294352						
VERSION	A37240.1	GI:2294352					
KEYWORDS	unidentified.						
SOURCE	unclassified.						
ORGANISM	unclassified.						
REFERENCE	1 (bases 1 to 23)						
AUTHORS	Patker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.						
TITLE	EPICAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE OR PROTEIN KINASE AND ASSAYS USING THEM						
JOURNAL	Patent: WO 9403609-A 9-17-FEB-1994;						
COMMENT	IMP CANCER RES TECH (GB)						
FEATURES	Other Publication Jp 8503124T 960409.						
source	1..23						
BASE COUNT	7 a	2 c	4 g	5 t	5 others		
ORIGIN	1..23						
Query Match	89.6% Score 20.6; DB 5; Length 23;						
Best Local Similarity	100.0% Pred. No. 1.3;						
Matches	23: Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
QY 1 AARATGAGTGGATHTTTCAYAC 23							
Db 1 AARATGAGTGGATHTTTCAYAC 23							
RESULT 5	HSPH13K	3424 bp	mRNA	PRI	24-AUG-1995		
LOCUS	DEFINITION	H.sapiens mRNA for phosphatidylinositol 3-kinase.					
ACCESSION	U23090						
NID	9472990						

VERSION	229090.1	GI:472990
KEYWORDS	phosphatidylinositol 3-kinase.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
	Primates; Catarrhini; Hominoidea; Homo.	
REFERENCE	1 (bases 1 to 1068)	
AUTHORS	Volinia,S., Hiles,I., Ormondtroyd,E., Nizetic,D., Antonacci,R., Rocchi,M. and Waterfield,M.D.	
TITLE	Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene	
JOURNAL	Genomics 24 (3), 472-477 (1994)	
MEDLINE	95292146	
REFERENCE	2 (bases 1 to 3424)	
AUTHORS	Volinia,S.	
TITLE	Direct Submission	
JOURNAL	Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig Institute for Cancer, Research, 91 Riding House Street, London, W1P 8BT, UK	
FEATURES		
source	Location/Qualifiers	
	1..3424	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="3q25.3"	
	/cell_line="K562"	
	13..3219	
	/codon_start=1	
	/product="phosphatidylinositol 3-kinase"	
	/protein_id="CA82333.1"	
	/db_xref="PIR:g472991"	
	/db_xref="GI:472991"	
	/db_xref="SWISS-PROT:P42336"	
	/translation="MPRRSSGELMGJHLMPPRIIVLECLPNGMIVTLECREATLVYIKHLFKFARKYPLHOLLDESSYIFVSTQEAREREFEDTRLCRLRLOPKYVIEPVNREBEKILNREIFGALMPVCEPDMVADPEVODRRLNLTNCRAVDRLDLSRHSRAVYPRHYESSPELPHKTYNKLDRGQIVLVAVYVSPNNOKYTKLNHCYPERQVAEATIRKTRSMLSSEDLKLCVLEIQGYILKVCGEDEYETKPLSYKTIKRGIMLGRPNLKKMAKESLYSQLPMDCFMPSYSRISTATPMNGETSKSLMIVNRLRIKLICATYVNLINRDIIDYRTGIVTGHGEPICLDNVTQRPVCSNRMWMLNYDIYIPDLPRARLICSTISCVKRGKGAKEEHCPILAMGINLEDYTDITSGAKMLNLPVPHGLEDNLPGVGTSGNPKKTEPCLLEFEDMSSVYKPPDMVSTIEHAAVSVSRAGFSYSHAGLSRLARDNLRENDKEQLKAISTRPDLSEITDQEDFLMSHHKICYTIPILLIKLILSVKMSNDEVAQMYCLVKDMPPIKPPDAMLDLCNTPDPVNRGPAVNCLEKTYILDDKLSQYLIQIQLVYLKLYEQLDNLVFLFLKALTNORIGHFEFMHLKSEHANTVSOGRGLLESYCRACGMYLKLNRYEAMEKILNLDIOEKERDETOVKOMKFLVEOKMRPDMALOGILSPINPAHOLGNRLKECRIMSSAKRPILMNENPDMISLEFQNNBRILFNGEDLRDMETLQILIRIMENINOCGLDLEMLPYGCLISGDCVGLLEVNSHRTIMQIQCKGGLGALQFNSHTLHWLKKRNKGETIDNAIDLFTSCAGTCVATPFLIGTGMJOLKXVNLKDDGLFHIIDFGHFLDHRKKFEGYKREVPVLLTDFLIVLSKQGEOTKTRFERFERQCYKAVYLAIROHANTFLNLSFMUGSGMPELOSFDJAIIRKTLALDKRTDEALVEYPMKMDAHHGWTYMDWIFTHIKHALLN"	
BASE COUNT	1134 a	618 c
ORIGIN	709 g	963 t
Query Match	89.6%	Score 20.6; DB 10; Length 3424;
Best Local Similarity	78.3%	Pred. No. 2.3;
Matches	18; Conservative	5; Mismatches
	0; Indels	0; Gaps
	0;	
1 AARATGAYTGGAHTTTCAYAC	23	
11:11111111111111111111		
DB 3172 AAAATGATGGATCTCCAC	3194	
RESULT	6	
HSU79143	3207 bp	mRNA
LOCUS	Human phosphoinositide 3'-hydroxylase p110-alpha subunit mRNA,	
DEFINITION	complete cds.	
ACCESSION	U79143	
NID	91763625	
KEYWORDS	GI:1763625	

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
AUTHORS 1 (bases 1 to 3207)
STRIVANT,S.M., AHERN,J., CONROY,R.R., BARNETT,S.F., LEDDER,L.M., OLIFF,A. and HELMBROOK,D.C.
TITLE Catalytic Activity of the p110-alpha Subunit of Human Phosphoinositide 3'-Hydroxylase is Required for Signal Transduction
JOURNAL Bloorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Strivant,S.M., Aherm,J., Conroy,R.R., Barnett,S.F., Ledder,L.M., Oliff,A. and Helmbrook,D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs, Summeytown Pike, West Point, PA 19486, USA
FEATURES
source
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/product="phosphoinositide 3'-hydroxylase p110-alpha subunit"
/protein_id="AA39753.1"
/db_xref="pid:g1763626"
/db_xref="gi:1763626"
/translation="MPRSSGELGNGIHMPRIIVLECLPNGMIVTECLREATLIT IKHELFREARKYPLHQLDSESYTFVSTQAESEFEDEFERRLCDLRFPELSPH EVGNREERILNREIGFVIGMPVCEPDWVDFPRNINLVCKEADVLDLNSPH SRAMVYPPNVSSEPELPHRYIKKDKQITVIVIVISPNNDKQYTKLINDCYPE QVIAEIRKRTSMLSSEQLKCVLEQKYLKVGCGDEYFLERKYSQYKIRSC IMGRPNMLMAKESLYSQLPDCFTMPSYRISTATPYNGESTSISLVNIAL RIKILCATYVNNIRIDIKYVRIYHGEPLCDVNTQRPCSNPRNEMVNDIY IPDLPRAARCLCISGVKGRKAEHCPLANGINLFDYDTLVSGKALNMPVPH GLEDLNPGLGVGSSNPKETPCLEEDPMSVYKRPMSVIEHANSVSREAEFSY SHAGLSNRARDELRENDKEQKALSTRDPLSETTEDEKDFLMSRHVCVTIPILP KLLSVKMSRDEVAQMYCLVMDPPIKPEQAMELDCNPDVAVRGFAVRCLEKYL DDKLSQYLQVLYVLEQYLDMLVRFLEKALNQRIGHFEMHLSKSMHAKVTSO RFGLEISYCRACGMYLKHLNQVEAMEKILNIDILKQKDETKOVQKMFVEQMR RPDNDALOGFELPNPAPHQLGNLRLEECRINSARPLNEMNDISELFQNE IIFKNGDDLROMLTLQIIRIMENTWOGDLRMLPYGCLSTGCVGLLEVRSNT IMQIOCKGLKAGLQENSRHLHQLMDKNKGETYDAIDLFTSCAGYCAVTFILIG DRHNSINWAKDQGLFHDGFHLDRKKKFGYKRERVPVLDLFLVYSKAGQCT KTRERFERQEMCKRAYLAIRQANLFINLSMGLSGMPELOSFDIAYIRKTLADK TQGEALEYFKQNDAHGGMWTCKMDMIFHTIKQALN"

BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN

Query Match 89.6%; Score 20.6; DB 10; Length 3207;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGATGATGATTTTCAYAC 23
||:||||:||||:||||:||||:||||:
Db 3160 AARATGATGATGATTCACAC 3182

RESULT 7
LOCUS MMU03279 3207 bp mRNA ROD 25-MAY-1994
DEFINITION Mus musculus Balb/c phosphatidylinositol 3-kinase 110 kDa subunit
ACCESSION U03279
NID 9414994
VERSION U03279.1 GI:414994
KEYWORDS house mouse, house musculus
SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS 1 (bases 1 to 3207)
KIPPEL,A., ESCOBEDO,J.A., HIRANO,M. and WILLIAMS,L.T.
TITLE The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE 94167738
REFERENCE 2 (bases 1 to 3207)
AUTHORS Kippel-Giese,A.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1993) Anke Kippel-Giese, CVRI and Daichi Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
FEATURES
source
1..3207
/organism="Mus musculus"
/strain="Balb/C"
/db_xref="taxon:10090"
1..3207
/standard_name="p110 protein"
/codon_start=1
/product="phosphatidylinositol 3-kinase 110 kDa subunit"
/protein_id="AA18334.1"
/db_xref="pid:g414995"
/db_xref="gi:414995"
/translation="MPRSSGELGNGIHMPRIIVLECLPNGMIVTECLREATLIT IKHELFREARKYPLHQLDSESYTFVSTQAESEFEDEFERRLCDLRFPELSPH EVGNREERILNREIGFVIGMPVCEPDWVDFPRNINLVCKEADVLDLNSPH SRAMVYPPNVSSEPELPHRYIKKDKQITVIVIVISPNNDKQYTKLINDCYPE QVIAEIRKRTSMLSSEQLKCVLEQKYLKVGCGDEYFLERKYSQYKIRSC IMGRPNMLMAKESLYSQLPDCFTMPSYRISTATPYNGESTSISLVNIAL RIKILCATYVNNIRIDIKYVRIYHGEPLCDVNTQRPCSNPRNEMVNDIY IPDLPRAARCLCISGVKGRKAEHCPLANGINLFDYDTLVSGKALNMPVPH GLEDLNPGLGVGSSNPKETPCLEEDPMSVYKRPMSVIEHANSVSREAEFSY SHAGLSNRARDELRENDKEQKALSTRDPLSETTEDEKDFLMSRHVCVTIPILP KLLSVKMSRDEVAQMYCLVMDPPIKPEQAMELDCNPDVAVRGFAVRCLEKYL DDKLSQYLQVLYVLEQYLDMLVRFLEKALNQRIGHFEMHLSKSMHAKVTSO RFGLEISYCRACGMYLKHLNQVEAMEKILNIDILKQKDETKOVQKMFVEQMR RPDNDALOGFELPNPAPHQLGNLRLEECRINSARPLNEMNDISELFQNE IIFKNGDDLROMLTLQIIRIMENTWOGDLRMLPYGCLSTGCVGLLEVRSNT IMQIOCKGLKAGLQENSRHLHQLMDKNKGETYDAIDLFTSCAGYCAVTFILIG DRHNSINWAKDQGLFHDGFHLDRKKKFGYKRERVPVLDLFLVYSKAGQCT KTRERFERQEMCKRAYLAIRQANLFINLSMGLSGMPELOSFDIAYIRKTLADK TQGEALEYFKQNDAHGGMWTCKMDMIFHTIKQALN"

BASE COUNT 958 a 693 c 734 g 822 t
ORIGIN

Query Match 89.6%; Score 20.6; DB 12; Length 3207;
Best Local Similarity 78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGATGATGATTTTCAYAC 23
||:||||:||||:||||:||||:||||:
Db 3160 AARATGATGATGATTCACAC 3182

RESULT 8
LOCUS AF001075 3389 bp mRNA VRL 08-JUL-1997
DEFINITION Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic subunit fusion protein (gag-v-p3k) and retroviral env mRNAs, partial cds.
ACCESSION AF001075
NID 92245502
VERSION AF001075.1 GI:2245502
KEYWORDS Avian sarcoma virus 16, Avian sarcoma virus 16
SOURCE Viruses; Retroviral; Avian type C retroviruses.
REFERENCE 1 (bases 1 to 3389)
AUTHORS Chang,H.M., Aoki,M., Fruman,D., Auger,K.R., Bellacosa,A., Tsichlis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.

TITLE	TRANSFORMATION OF CHICKEN CELLS BY THE GENE ENCODING THE CATALYTIC SUBUNIT OF PI 3-KINASE
JOURNAL	Science 276 (5320), 1848-1850 (1997)
MEDLINE	97334438
REFERENCE	2 (bases 1 to 3389)
AUTHORS	Chang, H.M., Aoki, M. and Vogt, P.K.
TITLE	Direct Submission
JOURNAL	Submitted (23-APR-1997) Molecular & Experimental Medicine, The Scripps Research Institute, 10350 N. Torrey Pines Road, Mail-drop BCC-239, La Jolla, CA 92037, USA
FEATURES	Location/Qualifiers
source	1. .3389
gene	/organism="avian sarcoma virus 16"
	/db_xref="taxon:60629"
	<1. .3252
	/gene="gag-v-p3k"
	/note="fusion gene of gag and v-p3k"
	1. .78
misc-feature	/gene="gag-v-p3k"
	/note="derived from gag gene"
	<1. .3252
CDS	/gene="gag-v-p3k"
	/codon_start=1
	/product="gag-v-phosphoinositide 3-kinase catalytic subunit fusion protein"
	/protein_id="AAB62532.1"
	/db_xref="GI:2245503"
	/translation="QOLIRPASTLTTPGRIIKYVLDROKASHLMPRIIVLCEPLNG MYTLECREATLTITIKHELFKARKYPTQLODSSYFVSVTOEAEKEEFEDTR RCDLDFQPFKVIIEPVGNREKILNRKIGFIMGICFEDVAKTPEVDPERNLIN VCKEAYDLKRNAPHSRALVCPNPVSESLKHLTKNDKQIIVLVIVSPND KQYVTLKINDVCEQVIAIAIKTRMSLSSEQLCVLEVGKYLIVAVSCDEYL LKQYPTSOYKIRSCIMLGMPNIMLAKSLYTOPLDFTMPSSRSRSTAPYMN GEATKSLMTINSALIRITICATYVNVNIDIKIVRTIGTIGSGELCDNVTORVP CSNPRNMENISYMDYIPDIPRARLICISVSGRGAKKEHPLMNGNIMDYATIE IYSGKALNIMAYMPTGLIEDLNPITGVTSNGRGTCLTEHPMFNSPYEPMTYTE EHAANTISRELGNTSYAGLSNRIANDNEIREDKEDRLAICTRDPISLITBEKDFL WSHRHYCVTPEILPKLLSVKNSRDEVAQMCVLRWPIRPEQAMELDNYPDP MRAFAVRCLLEKYLTDKLSQYLIOVLYQVLYDNLVRLKALINQIHGF FHLKSEMHNKTVSQRFGLLSEYSCACGMVLAHSQVAMEKLINLTDILOEKDD ENQVOMKFLVEQMRPDEMDALOGFTSPNPAHOLGNLEECRIMSSAKRLIMNW ENPDIMSELLFQNNETIFKKNDDLRQDMLTQIRIMENINQNGDLRLPYGCSTI GDCVGLIEVYRSHITMOICKKGLGALDFNSHTLMKDNKKGEMDAIDLFTL SCAGYCAVATFIIGIDRHSNINWKDGGFHLDFHGLDHRKKRGYKRRVPEVLT ODLLIVISKQAOCTKREFEPOEMCHKAYLAIRQHANLFINLFSMMLSGSEIQS FDDIAYIRKTLADKTEQALEYFMKQMDAHHGVTTKMDWIFHTIKQHANL"
misc-feature	79. .84
	/gene="gag-v-p3k"
	/note="sequence of unknown origin"
	85. .3252
misc-feature	/gene="gag-v-p3k"
	/note="derived from c-p3k gene"
	3250. .3370
3' UTR	3371. .>3389
CDS	/codon_start=1
	/product="teloviral env"
	/protein_id="AAB62533.1"
	/db_xref="PID:g2245504"
	/db_xref="GI:2245504"
	/translation="MARNS"
BASE COUNT	1071 a 655 c 751 g 912 t
ORIGIN	
Query Match	89.6%; Score 20.6; DB 17; Length 3389;
Best Local Similarity	78.3%; Pred. No. 2.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;	
OY	1 AARATGAGYTGAGTHTTYCAVAC 23
	: : :
DB	3205 AAAATGAGCTGATCTTCACAC 3227
RESULT	9
SEU77778	LOCUS
DEFINITION	Staphylococcus epidermidis plasmid pTue32 putative ABC transporter subunits (epiG), (epiE), and (epiF), putative membrane protein (epiH), EpiI' (epiI') and EpiI'' (epiI'') genes, complete cds.
ACCESSION	NID
VERSION	92196512
KEYWORDS	077778.1 GI:2196512
SOURCE	Staphylococcus epidermidis.
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE	1 (bases 1 to 2600)
AUTHORS	Peschel, A. and Goetz, F.
TITLE	Analysis of the Staphylococcus epidermidis genes epiF, -E, and -G involved in epidermal immunity
JOURNAL	J. Bacteriol. 178 (2), 531-536 (1996)
MEDLINE	96134989
REFERENCE	2 (bases 2473 to 5292)
AUTHORS	Peschel, A., Schnell, N., Hille, M., Entian, K.D. and Goetz, F.
TITLE	Secretion of the lantibiotics epidermin and gallidermin: sequence analysis of the genes gdmT and gdmH, their influence on epidermin production and their regulation by EpiO
JOURNAL	Mol. Gen. Genet. 254 (3), 312-318 (1997)
MEDLINE	97294510
REFERENCE	3 (bases 1 to 5292)
AUTHORS	Peschel, A., Schnell, N., Hille, M., Entian, K.-D. and Goetz, F.
TITLE	Direct Submission
JOURNAL	Submitted (15-JUN-1996) Mikrobielle Genetik, Universitaet Tuebingen, Waldaeufer Strasse 70/8, Tuebingen 72076, Germany
COMMENT	On Jun 14, 1997 this sequence version replaced gi:1176398. Other Staphylococcus epidermidis plasmid pTue32 lantibiotic epidermin genes can be found in sequence with GenBank Accession Number X62386.
FEATURES	Location/Qualifiers
source	1. .5292
	/organism="Staphylococcus epidermidis"
	/strain="Tue3298"
	/db_xref="taxon:1282"
gene	/plasmid="pTue32"
	complement(313. .1005)
	/gene="epiG"
CDS	complement(313. .1005)
	/gene="epiG"
	/function="involved in conferring immunity towards the lantibiotic epidermin"
	/note="EpiG; contains integral membrane domain"
	/codon_start=1
	/transl_table=11
	/product="putative ABC transporter subunit"
	/protein_id="AAB61164.1"
	/db_xref="PID:g1176401"
	/db_xref="GI:1176401"
	/translation="MKNEKLVNLFKRTIELTTPPLTIFVIFINLYKNMNTS ISQYTSFNALSSLIATITVYVYHFEESIGRHNILGPKRLMAYATLYFFSVL FCLLISTNFIKMSQNLKLEFVIVSSFNIIILIFISLFTKISFMSVGVVIT IENVYGIYVLDGKSWYIPIYTSYTRTSMFINNSVPLLTITIVSVIITTSLL LVAKRSGRSVOD
	complement(995. .1759)
gene	/gene="epiE"
	complement(995. .1759)
CDS	/gene="epiE"
	/function="involved in conferring immunity towards the lantibiotic epidermin"
	/note="EpiE; contains integral membrane domain"
	/codon_start=1
	/transl_table=11
	/product="putative ABC transporter subunit"
	/protein_id="AAB61162.1"
	/db_xref="PID:g1176400"

```

/db_xref="GI:1176400"
/translation="MKETKAEVLRKFKNTATAYVFLSLPFLGFLALFALRQSSPK
DIDISFLSLHLEMPYFVWPPIITICACSSMIGIERNGLFNYSNNNSINKMRSKI
LIMSJALLHLMFLVALTGMMVLGSGMSLGLLILFLFVFLSLPFLGFLALFALRQSSPK
CEGVLSIFLIALTLISIIIFITMSKFWIPWFSIGRIPIITLSLNGTLMSSNS
YENDLSALATIVISIIIESIFEMSNKSNINK"
complement(1761..2456)
/gene="epif"
complement(1761..2456)
/gene="epif"
/function="involved in conferring immunity towards the
antibiotic epidermin"
/notes="Epif: contains ATP-binding domain"
/codon_start=1
/transl_table=11
/product="putative ABC transporter subunit"
/protein_id="AAB61163.1"
/db_xref="PID:g1176399"
/db_xref="GI:1176399"
/translation="MTYKIKETRNLRNKKFKETIILKDNILKAYPNVYGLGINGACK
STLKITIGILPQTSGLYLDGSPILRKLNLGSLSPPTYNLSADQMLKVALN
ENIDNEISKVELNLNVEPQKVKNSLGMKQRLGAMALIKRKLPLVDEPNSGL
DPYGLQLRELKSLTLNNTSYIISHSLSLQGLADHVGIIHNGKLEYOEKNTDEN
LEDVFNITKSGK"
2615..3607
/gene="epih"
2615..3607
/gene="epih"
/function="involved in epidermin secretion"
/notes="Epih"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAB61159.1"
/db_xref="PID:g2196513"
/db_xref="GI:2196513"
/translation="MYFLPLFMVLVIAIILFKKSHRNLIYVATQIVGLICFY
ILAVNFSIVILMEILSLILFLFNKSLIKRSGSKLSNLSGSLILL
WIYTPSLFINKSLAVIYVIFHTVSLTFLPSLISLFLSKRDKDISI
YIVAGSLIEDKVPPLQSRIDICKNWDKDKHTKTEFLSGGQKDESEQAMK
KYLHAGVGNENRILICENKSQSTYENLKFSKLSIKNATIIICSNYHILRAVFAKL
KYLHAGVGNENRILICENKSQSTYENLKFSKLSIKNATIIICSNYHILRAVFAKL
complement(3744..4751)
/gene="epit"
complement(3744..4751)
/gene="epit"
/function="involved in epidermin secretion"
/notes="3" end of the epit gene which is disrupted by a
frame shift mutation"
/codon_start=1
/transl_table=11
/product="Epit"
/protein_id="AAB61160.1"
/db_xref="PID:g2196514"
/db_xref="GI:2196514"
/translation="MTLVCTPLIIVIIPIGINMKRVSQSGAPAKISYYNSR
LSTIKLITLSTYNIKIKNTLKNIDIELHAKYLSFEPIMNLILFINITGILE
LGYILMENMKSGDMFAVLYLQIINPIVSTISYEVORAGISSDLIKINKEPE
EGLYIKTYNNVQMEINDNFTKDNQIINSISLDHAKYIINIGESCGSTEL
NILAGLNTGYNIYLDLDRKQSFYSRMLFSYITDQILDELTVANNIYINEN
ISIEIIONACKKTNLSFLQNKNSFSISPSDSINLSIGOKRLVLRAPLQKNLLE
Y"
complement(4759..5109)
/gene="epit"
complement(4759..5109)
/gene="epit"
/function="involved in epidermin secretion"
/notes="5" end of the epit gene which is disrupted by a
frame shift mutation"
/codon_start=1
/transl_table=11
/product="Epit"
/protein_id="AAB61161.1"

```

```

/db_xref="PID:g2196515"
/db_xref="GI:2196515"
/translation="MIGTSELYIPIILRNALNKSLSLTDKIVIIILICFGLTLFSGV
STYLGIQGIKIIIONISVWNKVIKIPYFHLKNSASNLTSRLVNDNTITRFSVE
FIFSYSTNFIYV"
BASE COUNT 1944 a 645 c 658 g 2045 t
ORIGIN

Query Match 80.9%; Score 18.6; DB 1; Length 5292;
Best Local Similarity 76.2%; Pred. No. 25;
Matches 16; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGAGTGGATHTTTCAY 21
||:||||:||||:||||:|
Db 3530 AAAATGATGGATATTCAT 3550

RESULT 10
LOCUS HS69B13 310617 bp DNA HTG 10-JUN-1999
DEFINITION Homo sapiens chromosome 6 clone 69B13, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL035698
VERSION AL035698.6 GI:5050992
KEYWORDS HTG; HTGS_PHRASEL.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 310617)
REFERENCE
AUTHORS Bates,K.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humuerysanger.ac.uk Clone requests: clonerequests@anger.ac.uk
On Jun 11, 1999 this sequence version replaced gi:4902683.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known. 800 n s separate
segments. Unfinished: dJ69B13 Contig_ID: 03436 acc-AL035698
length: 304718 bp Unfinished: dJ69B13 Contig_ID: 02949
acc-AL035698 length: 2867 bp Unfinished: dJ69B13 Contig_ID: 00022
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Location/Qualifiers
source 1..310617
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="69B13"
BASE COUNT 92374 a 57041 c 59286 g 100315 t 1601 others
ORIGIN

Query Match 78.3%; Score 18; DB 34; Length 310617;
Best Local Similarity 72.7%; Pred. No. 81;
Matches 16; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 1 AARATGAGTGGATHTTTCAY 22
||:||||:||||:||||:|
Db 268820 AAAATGAGTGGATATTCACA 268841

RESULT 11
DMU12634

```

LOCUS	DMU12634	5392 bp	mRNA	INV	20-JUL-1995
DEFINITION	Drosophila melanogaster tollold related-1 (tldr-1) mRNA, complete cds.				
ACCESSION	U12634				
NID	9563143				
VERSION	U12634.1		GI:563143		
KEYWORDS	fruit fly,				
SOURCE	Drosophila melanogaster				
ORGANISM	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyridroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 5392)				
AUTHORS	Nguyen,T., Jmal,J., Shimmel,M.J., Arora,K. and O'Connor,M.B.				
TITLE	Characterization of tollold-related-1: a BMP-1-like product that is required during larval and pupal stages of Drosophila development				
JOURNAL	Dev. Biol. 166 (2), 569-586 (1994)				
MEDLINE	95113176				
REFERENCE	2 (bases 1 to 5392)				
AUTHORS	O'Connor,M.B.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-JUL-1994) Michael B. O'Connor, University of California, Molecular Biology and Biochemistry, Irvine, CA 92717-3900, USA				
FEATURES	Location/Qualifiers				
Source	1..5392				
	/organism="Drosophila melanogaster"				
	/strain="Canton S"				
	/db_xref="taxon:7227"				
	/clone="pMB01460"				
	/clone_lib="pMB40 library of Nick Brown"				
	/map="96A"				
	/dev_stage="0-8 hour embryo"				
	345..4739				
gene	/gene="tok"				
	/note="tolkin"				
	/allele=""				
	/db_xref="Flybase:FBgn0004885"				
	345..4739				
CDS	/gene="tok"				
	/note="Tldr-1"				
	/codon_start=1				
	/evidence=experimental				
	/product="collold related-1"				
	/protein_id="AAI0057.1"				
	/db_xref="PID:9563144"				
	/db_xref="GI:563144"				
	/translation="MRRRGTPGVSTNCCVLLATGLVYLIVSHAKNPPHGGLP PAPVPASDVQVATENNAHREPPVDSVLOLRGSDAVDEDEDEGLKHON LEHQOFGAVLTGIRLRAPRPSKTYTEQMLRLVTPVPSGDIIMDPKAGFNGDIA LPEDSGPVVAHTIDDEEOQROEVLADLSPVHPDPAVSAGANSPKFNATFC LDLEKLOEYVHEGLQVEEGLTDIIRKTKTPSGPYNLKAGOSOSOGANPVRK GSESPVAAOASYSKNDVLPSPDRIHNDKARLTIAIGKSSNKQSLIRNPTRR IQSNKDIEDDSGNPFAERPSPQIVPSLEQGSTPAHYTSNAGNGMQLDDEYDF HSRQVRRRRHRGPKQRSNQLHEALQLRLELSVPIEGHHROSHTQQAHRH HGHSHRRKROOHHRRKQSTTOYNNHSPFNOQAQOOSVEMDLKATST DEEELPHRYAVATFAKERIMDYGVIPYINDSGIKALFTKLAMRWENSTCIKE ERDEIHNIVTFVTRSGCCSFVGRKNGPOALSIGNCKQFGIYVHEIGVGFWMF EHTRDPERKHVYIEHNNTMGODYNENLDEPDSICMAYDYSIMHANTSKG YLDTLPLEAKGKRPELIGORLRSQSDIAQNLKCYKPCGRTPDSSGGLFASPSH TAGLSNTEHCEWRITATVGERVELKLENNIRKSNKCEYDTEIRGVEKSPDLIG RFCKDVKEVIRITRESSRLIYVNTHRLEGFRKAEDDVYCGGELSDVAVGLRES NYPDYLVKCEVWKITVPESYVQALKEGSEVENHSDCYDVEYRDPQDAPLIG VFCQKPPKPNSSNSMYVKEFSDVTSQKGFSAVPEKEDDECTONHGGHEG LGGEGCGRIGFELSHDKKHEADCGGVITPNCTISPSPEYPIILKEITVYAR PKHISLNTFTFDELRTHOOSDCGYDSYVTSKIGENRKLRIITFCSSSIPRATSE SNALRLEHSGKSIOFSGFAVFTDIDECAVNNGGCQHECRNTIGSTICCHNGYSK HENHCKDEGCKEISAPFCTISPVPDPSYPNADVCWHITTPGHRILINFEF VESHOECTYDVAAYDGESSVLRGCGDKIFPISSTNOMYATDKNKNKNO FTASHSTACGGLRATSOVOOYFHARGNDQDGMDCENTIAAPDSYQLFELT DISENTEPDYVOVFSIDIDYGOAGVGCYGNVLPDINSMTSHLLYRFTDGS VPAKGSASTYAVNSGCTIDSDDEVNSYSSEMTTPPGSLKTIYIEDTDETFEIS DFSNQALTFNQYIGRYSILPNRYISGA"				

polyA_site	5392	/evidence=experimental
BASE COUNT	1378 a 1443 c 1460 g 1111 t	
ORIGIN		
Query Match	78.3%	Score 18; DB 36; Length 5392;
Best Local Similarity	72.7%	Pred. No. 51;
Matches 16; Conservative	5; Mismatches	1; Indels 0; Gaps 0;
Oy	2 ARATGATGATGATTTTCACAC 23	
Db	5269 AGATGATGATGATTTTCACAC 5290	
1: :		
RESULT 12		
LOCUS	DMU34777	5381 bp mRNA INV 25-APR-1996
DEFINITION	Drosophila melanogaster tolkin mRNA, complete cds.	
ACCESSION	U34777	
NID	g1002985	
VERSION	U34777.1	GI:1002985
KEYWORDS		
SOURCE	fruit fly.	
ORGANISM	Drosophila melanogaster	
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE	Finelli,A.L., Xie,T., Bossie,C.A., Blackman,R.K. and Padgett,R.W.	
AUTHORS	The tolkin gene is a tollold/BMP-1 homologue that is essential for Drosophila development	
TITLE	Genetics 141 (1), 271-281 (1995)	
JOURNAL	2 (bases 1 to 5381)	
MEDLINE	Finelli,A.L., Bossie,C.A. and Padgett,R.W.	
REFERENCE	Direct Submission	
AUTHORS	Submitted (23-AUG-1995) Richard W. Padgett, Waksman Institute,	
JOURNAL	Rutgers University, Hoes Lane, Piscataway, NJ 08855, USA	
FEATURES	Location/Qualifiers	
source	1..5381	
gene	/organism="Drosophila melanogaster"	
	/db_xref="taxon:7227"	
	/map="96A"	
	330..4724	
	/gene="tok"	
	/note="tolkin"	
	/allele=""	
	/db_xref="FlyBase:FBgn0004885"	
	330..4724	
	/gene="tok"	
	/note="tollold homolog; Method: conceptual translation	
	supplied by author."	
	/codon_start=1	
	/protein_id="AAC47015.1"	
	/db_xref="PID:g1002986"	
	/db_xref="GI:1002986"	
	/translation="MRRRRGPIGVSNTNCVILLATGLAVLISVHAKNRPBRLGLPA	
	PAPRPVASYDVQVATENAHRRPVPDVSYLQRLSGSPRAVDEDDVDEDEGLKRRH	
	LEHQQDFDAVIGTIRLRRLRRPRRSKTTLEQLRLRHVTKVSGDDIMDPCKAGGFMGDTLA	
	LPEDEKQGVAAHTPDIIDDEQROREVEGLADSPSPVADPAVSQAANSGRKFKFATQD	
	LDLEKLKQEVYHEGLQVEEGELTDIIRKTKRLPSSGPNVINKAGOSQOSQSANPAPTR	
	GSESPDYAASSYKNDVLPPESEORHTDPRPRTILAEIGPSKNTQSILHRVDTDK	
	ISQNRKIDDDSGNNFVAERRPQIVPTSLPEORSTPAHYSNAGSGMGLDDVDLFL	
	HSRQVRRRRRRGPKQKRSNOHLEHEFKLQRLRELSRPVTEGHHRSHQKQAHQKQ	
	LHQSHKKRRHQQKRRRQSTQTVANSPPFNOQAQOQSVEMDLKKEISK	
	DEEPLIRVAAYVAKKERIMDWGVIVYEIDGNFSIHKALFKLAMRWENSTQIKFV	
	DEDELRHPNYIVFVTRSGCCSPFGKRGNGPOLAISGRNCDKFGIVVHEGLVGVFNH	
	EHTPEDEKHHVIEHNHIMKGDYFNMLSPDEVDSGLMAVDYSIMHVARNTFSKGT	
	YLDIILPIEMKGRKREPIGORIRLSODIDQANLILYCPMGCRFPFSSGIFPSPSHY	
	TAGLSMETEHCERTITATYGERVELLENNMNTFSKNCETDIETIDGTFEKSPILIG	
	RFCKGVAKVETRESSMLLVYVTHVIEGRGKAEFDVCGGELSVDVAQVPLESP	
	NYPLDLYPNNECVARKITVPESYVQALTFQSEVENHSCVYDVEVDRGQDAPLLIG	

VCGYKPPPMKSSGNSMYKFPYSDNSVOKAGSAVEMKREVEDECETONHCECECTINT
 LGCEECSCRTGFEHSDKRCEDACGCVLETPNGTITSPEFPMPLKCEWEIYAP
 PKRHLNFTHEFDEGTARHQSDCGDVSIVYKLEENRKRIGTCGSSIPETATSE
 SNARLEFHEHSDKIORSGFAVFTDIDECVANNNGCQHECRNTISYICMCHNGISM
 HENGHCKDECEKYEISAPGCTIFSPNPDSPYKNDVCVWHFTTGHRIKLFNEFD
 VESHOCTYDNVAVDGESESSVILGRFGDKIPFISSTSNOMVWLKDKRKOKNG
 FTASHSTACGGYATATSOVQOEFYSHARFGODYDDMDCEMTTAAADNSYVOLIETP
 DISSENCFTFDYQVATSDIDYVGYGPMRGYCGCVLRODINSMTSHSLVLRKTDGS
 VPMKGFSASTVAVPNSGEYDHSDEVDENSTSSMVPFPGSLSTIETDTEETDEYS
 DFSGNOLVFNGOYRGYSLPKRYSGKA"

BASE COUNT 1388 a 1430 c 1457 g 1106 t
 ORIGIN
 Query Match 78.3%; Score 18; DB 37; Length 5381;
 Best Local Similarity 72.7%; Pred. No. 51;
 Matches 16; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

RESULT 13
 AC007744/c
 LOCUS AC007744 199046 bp DNA HTG 05-JUN-1999
 DEFINITION Homo sapiens clone NH0482H16, WORKING DRAFT SEQUENCE, 1 unordered
 pieces.
 AC007744
 MID 95001488
 VERSION AC007744.1 GI:5001488
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 199046)
 AUTHORS Waterston, R.H.
 JOURNAL The sequence of Homo sapiens clone
 unpublished
 REFERENCE 2 (bases 1 to 199046)
 AUTHORS Waterston, R.H.
 JOURNAL Direct Submission
 Submitted (05-JUN-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source 1.199046 contig of 199046 bp in length.
 location/Qualifiers
 1..199046
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="NH0482H16"
 BASE COUNT 58842 a 36351 c 38451 g 65402 t
 ORIGIN

Query Match 77.4%; Score 17.8; DB 35; Length 199046;
 Best Local Similarity 73.9%; Pred. No. 97;
 Matches 17; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 AARATGATGATGATTTTCAYAC 23
 DB 54724 AAGATGATGATTTTCAYAC 54702

RESULT 14
 AC003092/c
 LOCUS AC003092 108878 bp DNA PRI 06-NOV-1997
 DEFINITION Human BAC clone RG335022 from 7q21-q22, complete sequence.
 AC003092
 MID 92588611
 VERSION AC003092.1 GI:2588611
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 108878)
 AUTHORS Lacy, M and Harper, M.
 JOURNAL The sequence of H. sapiens BAC clone RG335022
 Unpublished (1997)
 REFERENCE 2 (bases 1 to 108878)
 AUTHORS Waterston, R.
 JOURNAL Direct Submission
 Submitted (06-NOV-1997) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 SUBMITTED BY:
 Genome Sequencing Center
 Department of Genetics
 Washington University
 St. Louis MO 63108, USA
 http://genome.wustl.edu/gsc
 mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this
 clone. It may be shorter because we only sequence overlapping
 clone sections once, or longer because we provide a small overlap
 between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate
 chemistry; an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one subclone; and the assembly was
 confirmed by restriction digest.

MAPPING INFORMATION:
 The sequence of this clone was established as part of a mapping and
 sequencing collaboration between the NHGRI Chromosome 7 Mapping
 Project (Eric D. Green, Director), John D. McPherson in the
 Department of Genetics (Washington University), and the Washington
 University Genome Sequencing Center. For additional information
 about the map position of this sequence, see
 http://www.nhgri.nih.gov/DIR/CTB/CHR7 or send
 mailto:egreen@nhgri.nih.gov

SOURCE INFORMATION:
 Clone RG335022 is from a release of the human BAC library
 CTRB-HS-A. The library contains cloned DNA from human sperm. See:
 Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J.
 Kim et al., Genomics 34:213-8 (1996). The clone is available from
 Research Genetics, Inc. (http://www.resgen.com).
 VECTOR: pBeloBAC11
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
 The actual start of this clone is at base position 1 of RG335022;
 the actual end is at base position 108878 of RG335022.

This clone contains STS SWSS3514 (NID:q1222801).
 location/Qualifiers

FEATURES
 source 1.108878
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="7"
 /clone="RG335022"
 /clone_1kb="CTRB-HS-A"

```

repeat_region      /map="7921-q22"
                    563..794
                    /rpt_family="L1"
repeat_region      1243..1272
                    /rpt_family="L1"
repeat_region      6077..6102
                    /rpt_family="L1"
repeat_region      13493..13826
                    /rpt_family="L1"
repeat_region      15133..15424
                    /rpt_family="ALU"
repeat_region      20066..20485
                    /rpt_family="L1"
repeat_region      25026..25305
                    /rpt_family="ALU"
misc_feature        29019..29203
                    /note="match to EST D20877 (NID:9504697)"
repeat_region      36139..36166
                    /rpt_family="L1"
repeat_region      37435..37522
                    /rpt_family="ALU"
repeat_region      37570..37667
                    /rpt_family="ALU"
repeat_region      37708
                    /rpt_family="ALU"
repeat_region      38687
                    /rpt_family="ALU"
repeat_region      40727
                    /rpt_family="ALU"
repeat_region      42843..42878
                    /rpt_family="L1"
misc_feature        43587..43791
                    /note="match to EST D20858 (NID:9504678)"
repeat_region      45232..45303
                    /rpt_family="PTR"
repeat_region      50142..50216
                    /rpt_family="PTR"
repeat_region      51297
                    /rpt_family="L1"
repeat_region      60928..61221
                    /rpt_family="ALU"
repeat_region      62463..62800
                    /rpt_family="L1"
repeat_region      65195..66897
                    /rpt_family="PTR"
repeat_region      65222..66897
                    /rpt_family="PTR"
repeat_region      71038..71064
                    /rpt_family="L1"
repeat_region      75188..75478
                    /rpt_family="ALU"
repeat_region      81096..81385
                    /rpt_family="ALU"
repeat_region      82157..82212
                    /rpt_family="L1"
repeat_region      92199..92490
                    /rpt_family="ALU"
repeat_region      95111..95142
                    /rpt_family="L1"
repeat_region      95520..95820
                    /rpt_family="ALU"
repeat_region      96048..96393
                    /rpt_family="MER"
repeat_region      96861..96929
                    /rpt_family="ALU"
repeat_region      96932..96977
                    /rpt_family="L1"
repeat_region      102621
                    /rpt_family="L1"
repeat_region      104717..104998
                    /rpt_family="ALU"

```

```

BASE COUNT      34576 a 20570 c 20710 g 33022 t
ORIGIN
Query Match      75.7% Score 17.4: DB 11: Length 108878:
Best Local Similarity 69.6%: Pred 1.4e+02:
Matches 16: Conservative 5: Mismatches 2: Indels 0: Gaps 0:
QY      1 AARATGGAYTGATHTTTCAYAC 23
Db 56039 AAAATGGATTCACCTTCATAC 56017
|||||:|||||:|||||
RESULT 15
LOCUS HUAC004605/259474 bp DNA PRI 24-JUL-1998
DEFINITION Homo sapiens Chromosome 16 BAC clone C19875K-A-248F7, complete
sequence.
ACCESSION AC004605
VERSION 93337395
KEYWORDS AC004605.1 GI:3337395
SOURCE HTG.
ORGANISM human.
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 259474)
Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
Mason,T.M., Brandon,R., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
Homo sapiens Chromosome 16 BAC clone C19875K-A-248F7
Unpublished
2 (bases 1 to 259474)
Adams,M.D. and Loftus,B.J.
Direct Submission
Submitted (23-APR-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, Email:
bjoftus@tigr.org
3 (bases 1 to 259474)
Adams,M.D. and Loftus,B.J.
Direct Submission
Submitted (24-JUL-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jul 24, 1998 this sequence version replaced gi:3228501.
Address all correspondence to: Mark Adams The Institute for Genomic
Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail
address: humgent@tigr.org. The orientation of the sequence is from
sp6 end to t7 end. Genes were identified by a combination of five
methods including: XGRATL (available by anonymous ftp from
artbur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), GenScan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCANW.html) searches of the
complete sequence against a peptide database, and the Human gene
index database at TIGR (http://www.tigr.org/tib/tgi.html).
Genes without peptide homology having spliced EST hits are termed
'Unknown gene product'. Genes encoding tRNAs are predicted by
tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
FEATURES
Source
1..259474
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q22-q23"
/clone="A-248F7"
17061..17175
/note="32160, SHGC-18227, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G19282"
46411..46528
/note="7608, STS1-csRL-24g1-uA/csRL-24g1-uZ, Chr. -, Homo
sapiens"
/db_xref="dbSTS:G02122"
71868..71999
/note="32826, A002D07, Chr. -, Homo sapiens"
/db_xref="dbSTS:G19948"
STS

```

STS 72589. 72759
 /note="42758, D1S1053, Chr. -, Homo sapiens"
 /db_xref="dbSTS:628812"
 BASE COUNT 74059 a 49679 c 51651 g 84031 t 54 others
 ORIGIN

Query Match 75.7%; Score 17.4; DB 11; Length 259474;
 Best Local Similarity 69.6%; Pred. No. 1.6e+02;
 Matches 16; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 AARATGAGTGTGATTTTCAYAC 23
 ||: ||| :||||: ||: ||:
 Db 226583 AACTGGGCTGATATTTCACAC 226561

Search completed: September 13, 1999, 15:55:13
 Job time: 4541 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 14:39:33; Search time 255.05 Seconds

(without alignments)
22.562 Million cell updates/sec

Title: US-09-325-095-12

Perfect score: 23

Sequence: 1 AARATGATGATGATHTTCATAC 23

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20.6	89.6	3412	1	Human p110 cDNA. R
2	20.6	89.6	3207	1	p110 cDNA. Recombi
3	20.6	89.6	23	1	Primer for p110 CD
4	20.6	89.6	3498	1	Ptdins 3-Kinase 11
5	20.6	89.6	23	1	Ptdins 3-Kinase pr
6	16.4	71.3	40875	1	Insert from cosmid
7	16.4	69.6	381	1	H. pylori secreted
8	16.4	69.6	516	1	H. pylori secreted
9	16.4	69.6	800	1	MART-1 antigenic p
10	16.4	69.6	339	1	H. pylori GHPD 669
11	15.8	68.7	400	1	Staphylococcus aur
12	15.8	68.7	1959	1	Staphylococcus faec
13	15.8	68.7	5855	1	Staphylococcus faec
14	15.8	68.7	488	1	Enterococcus faec
15	15.8	68.7	3969	1	Bell methylase gen
16	15.6	67.8	3969	1	Chicken protein ty
17	15.6	67.8	3969	1	Chicken transmembr
18	15.4	67.0	5113	1	Continuation (3 of
19	15.4	67.0	1701	1	Human tyrosine dep
20	15.4	67.0	4752	1	Mouse p53 binding
21	15.4	67.0	4815	1	Nucleotide sequenc
22	15.4	67.0	4799	1	Nucleotide sequenc
23	15.4	67.0	1497	1	H. pylori cell env
24	15.4	67.0	1587	1	H. pylori cell env
25	15.4	67.0	1513	1	Streptococcus pneu
26	15.4	67.0	6186	1	Streptococcus pneu
27	15.4	67.0	1524	1	DNA encoding a S.
28	15.4	67.0	874	1	Clathrin-associat
29	15.4	67.0	6835	1	Enterococcus faec
30	15.4	67.0	7947	1	Enterococcus faec
31	15.4	67.0	2111	1	H. pylori GHPD 103
32	15.4	67.0	1687	1	H. pylori GHPD 170
33	15.4	67.0	3252	1	Maize phosphatidyl
34	15.2	66.1	4157	1	MAGE-2 gene. Nucle
35	15.2	66.1	1640	1	MAGE-3 cDNA. Nucle
36	15.2	66.1	2226	1	MAGE-5 cDNA. Nucle
37	15.2	66.1	2305	1	MAGE-5 genomic DN
38	15.2	66.1	1844	1	Parainfluenza viru
39	15.2	66.1	4157	1	Tumour rejection a
40	15.2	66.1	1640	1	Tumour rejection a
41	15.2	66.1	2226	1	Tumour rejection a
42	15.2	66.1	2305	1	Tumour rejection a
43	15.2	66.1	998	1	Joined p1cr2, whic

ALIGNMENTS

C	44	15.2	66.1	2305	1	T01165	MAGE-51 gene, Dete
C	45	15.2	66.1	19932	1	T46159	Cagl locus, Helico
ALIGNMENTS							
RESULT 1							
ID	051156	standard; cDNA; 3412 BP.					
AC	051156;	(first entry)					
DT	12-APR-1994						
DE	Human p110 cDNA.						
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;						
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;						
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;						
OS	Human.						
FM	Key	Location/Qualifiers					
FT	cds	1..3207					
FT		/*tag= a					
FT		/note= "PI3- kinase p110"					
DR	W09321328-A.						
PT	28-OCT-1993.						
PR	13-APR-1993; G00761.						
PA	(LUDW-) LODWIG INST CANCER RES.						
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;						
PI	Parker PJ, Volinia S, Waterfield MD;						
DR	WPI: 93-351738/44.						
DR	P-PSDB: R43342.						
PT	Recombinant polypeptide(s) - with phosphoinositide-3 kinase						
PT	activity, useful for controlling cell proliferation						
PS	Claim 7; Fig 16; 146pp; English.						
CC	Southern blot analysis was performed using a bovine cDNA probe contg.						
CC	a fragment of a PI3-kinase-encoding sequence and human cDNA isolated						
CC	from a cDNA library constructed from mRNA isolated from the human						
CC	cell line K562. Positive clones were sequenced to give the human						
CC	PI3 kinase p110 sequence shown. This sequence has 95 percent						
CC	homology with the bovine sequence. The domain encoding residues 19-						
CC	100 of human p110 is sufficient to encode the kinase which will						
CC	associate with the p85 kinase subunit. The gene may be used to						
CC	provide a protein with PI3 kinase activity, and is useful for						
CC	screening for (ant)agonists of PI3 kinase activity which could be						
CC	useful for stimulation or inhibition of cell proliferation and hence						
CC	prophylaxis or therapy. Platelet or neutrophil activity or blood						
CC	glucose levels can be controlled using the kinase.						
CC	See also 051155 and 057522-3.						
SQ	Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;						
Query Match							
Best Local Similarity		89.6%;	Score 20.6;	DB 1;	Length 3412;		
Matches 18; Conservative		78.3%;	Pred. No. 0.32;				
		5;	Mismatches 0;	Indels 0;	Gaps 0;		
QY	1 AARATGATGATGATHTTCATAC 23						
DB	3160 AARATGATGATGATCTCCACAC 3182						
RESULT 2							
ID	051155	standard; cDNA; 3207 BP.					
AC	051155;						
DT	12-APR-1994	(first entry)					
DE	p110 cDNA.						
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;						
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;						
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;						
SS							
FM	Key	Location/Qualifiers					
FT	cds	1..3207					
FT		/*tag= a					

FT /note- "P13- kinase p110"
 PN MO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI ParkerPJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB; R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a P13-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC P13 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with P13 kinase activity, and is useful for
 CC screening for (ant)agonists of P13 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC for prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 SQ Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 89.6%; Score 20.6; DB 1; Length 3207;
 Best Local Similarity 78.3%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;
 Matches 18; Conservative 5;

1 AARATGAYTGATHTTTCAYAC 23
 3160 AARATGATGATGATCTCCACAC 3182

RESULT 3
 059012
 ID 059012 standard; DNA: 23 BP.
 AC 059012;
 DT 12-APR-1994 (first entry)
 DE Primer for p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification.
 OS Synthetic.
 PN MO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI ParkerPJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 38; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. The
 CC unamplified library was plated on E. coli K12 PLF-F and screened
 CC with the labelled primer shown and the primer of 059013. Hybridising
 CC clones were sequenced.
 CC See also 051156, 059012-23 and 057522-3.
 SQ Sequence 23 BP; 7 A; 2 C; 4 G; 5 T;

Query Match 89.6%; Score 20.6; DB 1; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.14;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AARATGAYTGATHTTTCAYAC 23
 DB 1 AARATGATGATGATCTCCACAC 23

RESULT 4
 057012
 ID 057012 standard; cDNA to mRNA; 3498 BP.
 AC 057012;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 KD catalytic subunit cDNA.
 KW 110 KD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /product= p110
 PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB; R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 KD catalytic subunit of the phosphatidylinositol (Ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the Ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SQ Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 89.6%; Score 20.6; DB 1; Length 3498;
 Best Local Similarity 78.3%; Pred. No. 0.33; Mismatches 0; Indels 0; Gaps 0;
 Matches 18; Conservative 5;

1 AARATGAYTGATHTTTCAYAC 23
 3160 AARATGATGATGATCTCCACAC 3182

RESULT 5
 057018
 ID 057018 standard; DNA: 23 BP.
 AC 057018;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase probe #1.
 KW 110 KD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.

PA (INNO-) INNOGENETICS NV..
 PI Scarce T, Van Broekhoven A;
 DR WPI: 97-132645/12.
 CC New nucleic acid encoding *Saccharomyces pectinase* - useful in food
 CC processing, also its regulatory sequences are useful for expression
 PT of heterologous protein
 PS Example 1: Fig 2: 65pp; English.
 CC This sequence represents the insert from cosmid 109, containing
 CC *Saccharomyces cerevisiae* strain FY1679 DNA. This sequence contained the
 CC pectinase gene of the invention (see 180042). Pectinases are synthesised
 CC by plants and a variety of microorganisms, the best characterised
 CC pectinase being polygalacturonase. Pectinases play roles in the invasion
 CC of plant tissues by phytopathogens, the spoilage of fruits and
 CC vegetables, and also in their food processing and plant biological
 CC applications. Vectors containing the pectinase coding sequence are used
 CC for production of homologous or heterologous proteins in transformed host
 CC cells. The pectinase sequence may include regulatory and/or coding
 CC sequences, including the sequence encoding the pectinase signal peptide
 CC (which can be used in any recombinant system). The pectinase is useful in
 CC food processing (e.g. preparation of fruit juices), typical heterologous
 CC proteins that can be made are (poly)peptides of bacterial, animal or
 CC viral origin. Also, probes derived from the pectinase coding sequence can
 CC be used to detect the coding sequence and primers for amplification are
 CC used to isolate related genes.
 SO Sequence 40875 BP; 12953 A; 7579 C; 8292 G; 12051 T;

Query Match 71.3%; Score 16.4; DB 1; Length 40875;
 Best Local Similarity 68.2%; Pred. No. 59;
 Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 AARATGATGATGATHTTCAYCA 22
 DB 30251 AARATGACTGATCTTCATA 30272

RESULT 7
 T67697 standard; DNA: 381 BP.
 AC T67697;
 DT 15-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein ORF 5869090.aa.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
 OS Helicobacter pylori.
 FS key Location/Qualifiers
 FT cds 1..381
 FT /*tag= a
 FT /note= "no stop codon given"
 PN MO9640893-A1.
 PD 19-DEC-1996.
 PE 06-JUN-1996; 009122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR P-PSDB: W20547.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 23; Page 308; 1481pp; English.
 CC This sequence encodes a H. pylori secreted or periplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant

CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 SO Sequence 381 BP; 133 A; 59 C; 71 G; 118 T;

Query Match 69.6%; Score 16; DB 1; Length 381;
 Best Local Similarity 77.8%; Pred. No. 41;
 Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGGATGATGATHTTCAYCA 22
 DB 60 TGGATGATGATTCACA 77

RESULT 8
 T67862 standard; DNA: 516 BP.
 AC T67862;
 DT 14-JUL-1997 (first entry)
 DE H. pylori secreted or periplasmic protein ORF 019p11016orf14.
 DE Vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
 KW secreted; periplasmic; ds.
 OS Helicobacter pylori.
 FS key Location/Qualifiers
 FT cds 1..516
 FT /*tag= a
 FT /note= "no stop codon given"
 PN MO9640893-A1.
 PD 19-DEC-1996.
 PE 06-JUN-1996; 009122.
 PR 07-JUN-1995; US-487032.
 PR 01-APR-1996; US-630405.
 PA (ASTR) ASTRA AB.
 PI Berglindh OT, Smith D, Mellgaard BL;
 DR WPI: 97-052306/05.
 DR P-PSDB: W20609.
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 PS Claim 23; Pages 753-754; 1481pp; English.
 CC The present sequence encodes a Helicobacter pylori secreted or
 CC periplasmic protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the
 CC bacterial DNA. The sequences were analysed for ORF of at least 180
 CC nucleotides, and the predicted coding regions defined by computer
 CC evaluation. To identify likely H. pylori antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences
 CC of interest, particular regions can be isolated from H. pylori by
 CC PCR amplification for recombinant polypeptide production, e.g. in
 CC E. coli hosts.
 SO Sequence 516 BP; 184 A; 69 C; 106 G; 157 T;

Query Match 69.6%; Score 16; DB 1; Length 516;
 Best Local Similarity 77.8%; Pred. No. 43;
 Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGGATGATGATHTTCAYCA 22
 DB 63 TGGATGATGATTCACA 80

RESULT 9

T78769
ID T78769 standard; cDNA; 800 BP.
AC T78769;
DT 02-MAR-1998 (first entry)
DE MART-1 antigenic protein nucleic acid melanoma antigen.
KW MART-1 antigen; melanoma; dendritic cell; T cell; recombinant;
KM haematopoietic stem cell; antigenic peptide; cancer; MHC;
OS major histocompatibility complex; ss.
OS Mus musculus.
PN W09729183-A2.
PD 14-AUG-1997.
PF 07-FEB-1997; U02063.
PR 08-FEB-1996; US-011433.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Hwu P, Reeves M, Rosenberg SA;
DR WPI: 97-415339/38.
PT New transduced dendritic cells for activating T cells - produced by
PT transducing haematopoietic stem cells with nucleic acid and
PT differentiating the cells into dendritic cells
PS Disclosure; Fig 3; 63pp; English.
CC A novel method has been developed for making a dendritic cell transduced
CC with a selected nucleic acid. The method comprises: (a) transducing a
CC haematopoietic stem cell (HSC) with the selected nucleic acid; and (b)
CC differentiating the transduced stem cell into a dendritic cell, thereby
CC making a dendritic cell transduced with the selected nucleic acid. The
CC present sequence represents the nucleic acid sequence for MART-1 which
CC encodes an antigenic protein, a major histocompatibility complex class
CC I bound melanoma antigen. The recombinant dendritic cells expressing
CC antigenic peptides can activate T-cells against target cells expressing
CC selected antigens in vivo. They can also be used for detecting T cell
CC mediated anti-cancer cell activity of a target antigenic peptide and
CC for altering the MHC class II antigenic repertoire of a dendritic cell.
CC They can be used for the treatment of cancers and cellular infections
CC and in diagnostic and cell screening assays. T cells activated by
CC contact with dendritic cells may be used to kill a target cell,
CC especially a cancer cell. The recombinant dendritic cells provide
CC broader antigen expression and more efficient MHC class I and class II
CC peptide loading, and the ability to expand the population of desired
CC dendritic cells, e.g. in culture. 175 C; 205 G; 191 T;
SQ Sequence 800 BP; 185 A; 175 C; 205 G; 191 T;

Query Match 69.6%; Score 16; DB 1; Length 800;
Best Local Similarity 63.6%; Pred. No. 46;
Matches 14; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

2 AGATGCAVTTGGATHTTTCAYAC 23
1:||||:||||:||||:||||:
Db 195 AGATGCAVTTGGATHTTTCCTAC 216

RESULT 10
X14176/c
X14176 standard; DNA; 339 BP.
AC X14176;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPD 669 gene.
KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease; ss.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 1..339
FT /tag= a
FT WO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HOMA-) HUMAN GENOME SCI INC.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-542293/46.

DR P-PSDB: W98457.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 1; Page 804; 2054pp; English.
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPD protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 339 BP; 112 A; 80 C; 48 G; 99 T;

Query Match 69.6%; Score 16; DB 1; Length 339;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 14; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 TGGATGGATHTTTCAYA 22
||||:||||:||||:||||:
Db 147 TGGATGGATHTTTCACA 130

RESULT 11
V75319/c
V75319 standard; DNA; 400 BP.
AC V75319;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus confg SEQ ID #1008.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HOMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 1791; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC S.aureus can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 400 BP; 145 A; 72 C; 55 G; 123 T;

Query Match 68.7%; Score 15.8; DB 1; Length 400;
Best Local Similarity 57.1%; Pred. No. 52;
Matches 12; Conservative 8; Mismatches 1; Indels 0; Gaps 0;

RESULT 15

V05762

ID V05762 standard; cDNA; 3969 BP.

AC V05762;

DT 12-MAY-1998 (first entry)

DE Chicken protein tyrosine phosphatase-lambda gene.

KM Chicken: protein tyrosine phosphatase-lambda; brain; probe; primer; PCR;

KW hybridisation; human CD45; amplification; alternative splicing product;

KM diagnosis; cancer; immune system disorder; ligand; ss.

OS Gallus domesticus.

FH Key Location/Qualifiers

FT CDS 91..3804

TT /*tag= a

/product= "protein tyrosine phosphatase

US5693488-A.

PD 02-DEC-1997.

PF 12-MAY-1994; 241853.

PK 12-MAY-1994; US-241853.

PA (UVR0) UNIV ROCKEFELLER.

PI Fang KS, Hanafusa H;

DR WPI; 98-031746/03.

DR P-PSDB; W44729.

DT Nucleic acid molecule encoding chicken protein tyrosine phosphatase

PT - specifically transmembrane protein tyrosine phosphatase-lambda,

PS Claim 2; Column 31-40; 51pp; English.

CC This is the nucleotide sequence which encodes the novel chicken protein

CC tyrosine phosphatase (PTP)-lambda. The sequence was isolated from a

CC chicken brain cDNA library, using, as a probe, a fragment encoding the

CC intracellular domain of the human CD45 sequence. The sequence was then

CC amplified using primers V05763-V05766. The gene has a transcript of

CC around 5.6 kb and is abundant in spleen and intestine. The sequence has

CC been shown to contain 5 alternative splicing products which vary near

CC the N-terminus of the encoded protein. Nucleic acid molecules,

CC especially encoding residues 22-1237, 22-509, 510-531 or 532-1237 of the

CC 1237 residue PTP-lambda protein, can be used as probes and primers to

CC detect levels of phosphatase expression. This is useful for the

CC diagnosis and treatment of diseases such as cancer or immune system

CC functional disorders. The protein can also be used to isolate ligands

CC of the PTP.

SQ Sequence 3969 BP; 1317 A; 816 C; 847 G; 989 T;

Query Match

Best Local Similarity 67.88; Score 15.6; DB 1; Length 3969;

Matches 15; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 RATGATGATGATTTTCAYA 22

Db 3265 AATGATGATGATTCAGA 3284

Search completed: September 13, 1999, 15:59:29
Job time: 4796 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 14:38:32 ; Search time 1694.61 Seconds
(without alignments)
26.772 Million cell updates/sec

Title: US-09-325-095-12

Perfect score: 23
Sequence: 1 AARATGATGATHTTCATAC 23

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20.6	89.6	292	23	D56232	D56232 HOM420G03B
2	20.6	89.6	264	24	N23534	N23534 YV91G08..r1
3	20.6	89.6	475	25	N44819	N44819 YV39A01..r1
4	20.6	89.6	405	30	AA237307	AA237307 mw96h11.r
5	20.6	89.6	331	33	AA445047	AA445047 v957h11.r
6	20.6	89.6	206	39	AA874470	AA874470 vx80F10.r
7	20.6	89.6	447	42	A1155136	A1155136 u459f06.r
8	20.6	89.6	544	46	A1411730	A1411730 EST240024
C 9	20.6	89.6	438	47	A1502906	A1502906 UI-R-C1-K
C 10	20.6	89.6	431	48	A1547814	A1547814 UI-R-C3-S
C 11	20.6	89.6	633	49	A1632172	A1632172 t859g07.x
12	20	87.0	453	45	A1324274	A1324274 mq72c02.y
13	19	82.6	433	31	AA298585	AA298585 EST114216
14	18	78.3	433	26	W76227	W76227 z4d4c12..r1
15	17.2	74.8	695	45	A1354431	A1354431 gu17c11.x
16	17	73.9	410	23	R83858	R83858 yb82e10.r1
17	17	73.9	593	30	AA212806	AA212806 mw84e10.r
18	17	73.9	530	30	AA239420	AA239420 mw98d08.r
19	17	73.9	394	35	AA575527	AA575527 vm03b08.r
20	17	73.9	441	42	A1156560	A1156560 ue43e06.r
21	16.4	71.3	465	22	R73756	R73756 Y155f02..r1
22	16.4	71.3	485	25	N44970	N44970 YV35g10..r1
C 23	16.4	71.3	410	33	AA443680	AA443680 zw86c09.s
C 24	16.4	71.3	326	35	AA585856	AA585856 28505 Lam
C 25	16.4	71.3	294	36	AA652512	AA652512 ns72a01.s
C 26	16.4	71.3	516	41	A1052103	A1052103 oy30a04.x
C 27	16.4	71.3	433	41	AU016499	AU016499 AU016499
28	16.4	71.3	584	41	AU016839	AU016839 AU016839
C 29	16.4	71.3	405	43	A1239854	A1239854 qh43h09.x
30	16.4	71.3	902	44	A1323676	A1323676 mq21g08.x
31	16.4	71.3	288	46	A1413704	A1413704 me03h10.x
32	16.4	71.3	503	47	A1523627	A1523627 t99sc06.x
33	16.4	71.3	243	48	A1605101	A1605101 vth04d12.x
C 34	16.4	71.3	471	50	A1692399	A1692399 wd68b03.x
C 35	16.4	71.3	560	50	A1692482	A1692482 wd70h12.x
C 36	16.4	71.3	206	50	AV029255	AV029255 AV029255
C 37	16.2	70.4	438	29	AA127013	AA127013 z116h06.s
38	16.2	70.4	525	30	AA260539	AA260539 vth06e07.r
C 39	16.2	70.4	430	33	AA427986	AA427986 zw50b01.r
40	16.2	70.4	349	37	AA727030	AA727030 vu36h05.r
C 41	16.2	70.4	439	43	A1187390	A1187390 qf31a02.x
C 42	16.2	70.4	474	47	A1464790	A1464790 mz81a05.y
C 43	16.2	70.4	436	47	A1485837	A1485837 EST244158
C 44	16.2	70.4	183	20	T27380	T27380 hbc2456 Hum
C 45	16	69.6				

ALIGNMENTS

RESULT 1
D56232/c D56232 292 bp mRNA EST 31-AUG-1995
LOCUS HOM420G03B Clontech human fetal brain polyA+ mRNA (#6535) Homo
DEFINITION sapiens cDNA clone GEN-420G03 5', mRNA sequence.
ACCESSION D56232
NID 9970756
VERSION D56232.1 GI:970756

NAME: EST.
 SOURCE: human.
 ORGANISM: Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (bases 1 to 292)
 Fujiiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
 Okuno, S., Ozaki, K., Shimizu, F., Shinada, Y., Shinomiya, H.,
 Takaiuchi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
 Mekawa, H., Shin, S. and Nakamura, Y.
 Fujiiwara et al. (1995)
 Unpublished (1995)
 TITLE
 JOURNAL
 COMMENT
 On Sep 21, 1992 this sequence version replaced g1:276155.

Contact: Tsutomu Fujiiwara
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawanuchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035
 Insert Length: 770 Std Error: 0.00
 High quality sequence stop: 331.
 Location/Qualifiers
 1..292
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_id="Clontech human fetal brain polyA+ mRNA
 (#6535)"

BASE COUNT 78 a 59 c 55 g 100 t
 ORIGIN

Query Match 89.6%; Score 20.6; DB 23; Length 292;
 Best Local Similarity 78.3%; Pred. No. 7.5;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Y 1 AARATGATGATGATTCACAC 23
 11:|||||:|||||:|||||:|
 Db 67 AAATGATGATGATTCACAC 45

RESULT 2
 N23534 264 bp mRNA EST 28-DEC-1995
 LOCUS Y91908.r1 Soares melanocyte 2NBHM Homo sapiens CDNA clone
 IMAGE:250142.5 similar to SP:P100_BOVIN P32871
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT ; mRNA sequence.
 ACCESSION N23534
 NID g1137684
 VERSION N23534.1 GI:1137684
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 264)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 TITLE
 JOURNAL
 COMMENT
 On Apr 14, 1993 this sequence version replaced g1:692929.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 High quality sequence stops: 225

Source: IMAGE Consortium, LNL.
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1135 Std Error: 0.00
 Seq primer: M13Rpl
 High quality sequence stop: 225.
 Location/Qualifiers
 1..264
 /organism="Homo sapiens"
 /db_xref="GDB:386784"
 /db_xref="taxon:9606"
 /map="14"
 /clone="IMAGE:250142"
 /clone_lib="Soares melanocyte 2NBHM"
 /sex="Male"
 /tissue_type="melanocyte"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: p773D (Pharmacia) with a modified
 polylinker. Site 1: Not I. Site 2: Eco RI. 1st strand CDNA
 was primed with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGAGCGCCGAGTCTTTTCTTTTCTTTT 3'],
 double-stranded CDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified p773 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldo. RNA from normal foreskin melanocytes
 (#5374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 94 a 53 c 52 g 60 t
 ORIGIN

Query Match 89.6%; Score 20.6; DB 24; Length 264;
 Best Local Similarity 78.3%; Pred. No. 7.5;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

Y 1 AARATGATGATGATTCACAC 23
 11:|||||:|||||:|||||:|
 Db 140 AAATGATGATGATTCACAC 162

RESULT 3
 N44819 475 bp mRNA EST 13-FEB-1996
 LOCUS Y939410.r1 Soares melanocyte 2NBHM Homo sapiens CDNA clone
 IMAGE:273594.5 similar to SW:P11A_BOVIN P32871
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
 mRNA sequence.
 ACCESSION N44819
 NID g1185985
 VERSION N44819.1 GI:1185985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 TITLE
 JOURNAL
 COMMENT
 On May 8, 1995 this sequence version replaced g1:800198.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: T7
High quality sequence stop: 402.
Location/Qualifiers

1..475

FEATURES
source
/organism="Homo sapiens"
/db_xref="GDB:3883236"
/db_xref="taxon:9606"
/clone_image="273594"
/clone_lib="Soares melanocyte 2NDHM"
/sex="Male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCGCACTTTTCTTTTCTTTT 3']; double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 170 a 90 c 85 g 128 t 2 others

ORIGIN

Query Match 89.6%; Score 20.6; DB 25; Length 475;
Best Local Similarity 78.3%; Pred. No. 7.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGATGATGATTCACAC 23

Db 201 AARATGATGATTCACAC 223

RESULT 4

LOCUS AA237307 405 bp mRNA EST 03-MAR-1997
DEFINITION m96h11.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:678597 5' similar to SW:P1LA_BOVIN_P32871 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.

ACCESSION AA237307
NID AA237307
VERSION AA237307.1 GI:1861328
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 405)
Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE *The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1404918.

FEATURES
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:418301
Putative full length read
vector to vector length is
Seq primer: -28m13 rev2 ET from Amersham.
Location/Qualifiers

source

1..405

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_image="678597"
/clone_lib="Soares mouse NML"
/tissue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCGCACTTTTCTTTTCTTTT 3']; double-stranded cDNA was ligated to Eco RI adapters (Pharmacia), digested with Not I and Eco RI sites of the modified pT73 vector. Library constructed and normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 141 a 79 c 90 g 95 t

ORIGIN

Query Match 89.6%; Score 20.6; DB 30; Length 405;
Best Local Similarity 78.3%; Pred. No. 7.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGATGATGATTCACAC 23

Db 25 AARATGATGATTCACAC 47

RESULT 5

LOCUS AA45047 331 bp mRNA EST 03-JUN-1997
DEFINITION v957h11.r1 Beddington mouse embryonic region Mus musculus cDNA clone IMAGE:865509 5' similar to SW:P1LA_MOUSE_P42337 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.

ACCESSION AA45047
NID 92157730
VERSION AA45047.1 GI:2157730
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 331)
Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE *The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 5, 1995 this sequence version replaced gi:798358.

FEATURES
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:509597
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 299.
Location/Qualifiers
1..331
/organism="Mus musculus"
/strain="C57BL/6 x DBA"
/db_xref="taxon:10090"
/map="21"
/clone_image="865509"

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="4"
/clone="IMAGE:181547"
/clone_lib="Scotres 2NDMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10"
note=Vector: pTR73D-Pac (Pharmacia) with a modified

```

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone IMAGE:1450211"
/clone-lib="Scares mouse uterus MMPu"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT7T3d-Pac (Pharmacia) with
a modified polylinker; 1st strand cDNA was prepared from
pregnant mouse uterus, and was then primed with a Not I - Eco
1950(47) primer. Double-stranded cDNA was ligated to Eco

```

RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 139 a 79 c 101 g 127 t 1 others
 Query Match 89.6%; Score 20.6; DB 42; Length 447;
 Best Local Similarity 78.3%; Pred. No. 7.3;
 Matches 18; Conservative 5; Mismatches 0; Gaps 0;
 Indels 0;
 OY 1 AARATGATGATGATTCATCAYC 23
 Db 403 AARATGATGATGATTCATCAYC 425

RESULT 8 A1411730 544 bp mRNA EST 09-FEB-1999
 LOCUS A1411730.1
 DEFINITION EST240024 Normalized rat placenta, Bento Soares Rattus sp. CDNA
 ACCESSION A1411730
 VERSION 1
 KEYWORDS A1411730.1 GI:4255234
 SOURCE EST.
 ORGANISM Rattus sp.
 Rattus sp. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 AUTHORS Lee, N.H., Glodet, A., Chandra, I., Mason, T.M., Quackenbush, J., Keriavage, A.R. and Adams, M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index
 JOURNAL Unpublished (1998)
 COMMENT On Apr 7, 1998 this sequence version replaced gi:3035706.
 Other ESTs: EST240023
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.

FEATURES
 source
 Location/Qualifiers
 1..544
 /organism="Rattus sp."
 /db_xref="taxon:10118"
 /clone="RPDJ40"
 /note="Normalized rat placenta, Bento Soares"
 /note="Organ: placenta; Vector: pRT3pac; Site_1: EcoRI; Site_2: NotI"

BASE COUNT 131 a 119 c 110 g 184 t
 ORIGIN

Query Match 89.6%; Score 20.6; DB 46; Length 544;
 Best Local Similarity 78.3%; Pred. No. 7.2;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AARATGATGATGATTCATCAYC 23
 Db 385 AARATGATGATGATTCATCAYC 363

RESULT 9 A1502906 438 bp mRNA EST 11-MAR-1999
 LOCUS A1502906
 DEFINITION UI-R-C1-Kn-b-09-0-UI.s1 UI-R-C1 Rattus norvegicus CDNA clone
 UI-R-C1-Kn-b-09-0-UI 3', mRNA sequence.
 ACCESSION A1502906

NTD 94400757
 VERSION A1502906.1 GI:4400757
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.
 TITLE 1 (bases 1 to 438)
 JOURNAL Normalization and subtraction: two approaches to facilitate gene discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT 97044477
 On Mar 16, 1998 this sequence version replaced gi:2961816.

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. CDNA Library Preparation: M.B. Soares Lab Clone distribution: Clones will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward.
 Location/Qualifiers
 1..438
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /map="7p14-15"
 /clone="UI-R-C1-Kn-b-09-0-UI"
 /clone_lib="UI-R-C1"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1 library is a subtracted library derived from the UI-R-C0 library, which is a subtracted library derived from the UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996)."

BASE COUNT 94 a 92 c 85 g 167 t
 ORIGIN

Query Match 89.6%; Score 20.6; DB 47; Length 438;
 Best Local Similarity 78.3%; Pred. No. 7.3;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AARATGATGATGATTCATCAYC 23

Db 399 AARATGACTGATCTTCACAC 377

RESULT 10
A1547814 431 bp MRNA EST 22-MAR-1999
UI-R-C3-sk-b-10-0-01.g1 UI-R-C3 Rattus norvegicus CDNA clone
DEFINITION
UI-R-C3-sk-b-10-0-01 3', mRNA sequence.
ACCESSION
A1547814
NID
94465302
VERSION
A1547814.1 GI:4465302
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 431)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
On Jun 5, 1998 this sequence version replaced gi:3187473.

FEATURES
source
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
*tel: 319 335 8250
Fax: 319 335 9565
Email: msocares@blue.weeg.uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the
normalized placenta library cDNA library preparation: M.B. Soares
Lab Clone distribution: clones will be available through Research
Genetics (www.resgen.com) The following repetitive elements were
found in this cDNA sequence: 1-70, >Poly_A\$Simple_repeat
Seq primer: M13 Forward.
Location/Qualifiers
1..431
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/map="1: 874H02; 7: 7421.3-7422.1; 4p16.3; 2"
/clone="UI-R-C3-sk-b-10-0-01"
/clone_lib="UI-R-C3"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; The UI-R-C3
library is a subtracted library of a series, ultimately
derived from a mixture of individually tagged normalized
libraries from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day
embryos, after a series of subtractions to reduce the
representation of cDNAs from which ESTs had already been
generated. The following serially subtracted libraries
were generated in this process: UI-R-C3, UI-R-C2p,
UI-R-C1, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of
3-5 nucleotides present between the Not I site and the
oligo-dT track which allows identification of the library
of origin of a clone within the mixture. The subtracted
library (UI-R-C3) was constructed as follows: pCIS amplified
cDNA inserts from UI-R-C2p clones from which 3' ESTs had
been derived was used as a driver in a hybridization with
the UI-R-C2p library in the form of single-stranded
circles. The remaining single-stranded circles (subtracted
library) was purified by hydroxapatite column
chromatography, converted to double-stranded circles and

electroporated into DH10B bacteria (Life Technologies) to
generate the UI-R-C3 library. This procedure has been
previously described (Bonaldo, Lennon and Soares, genome
Research 6:791-806, 1996)."
BASE COUNT 93 a 92 c 84 g 162 t
ORIGIN

Query Match 89.6%; Score 20.6; DB 48; Length 431;
Best Local Similarity 78.3%; Pred. No. 7.3;
Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
QY 1 AARATGACTGATCTTCACAC 23
Db 401 AARATGACTGATCTTCACAC 379

RESULT 11
A1632172c
LOCUS
A1632172c
DEFINITION
A1632172 633 bp MRNA EST 26-APR-1999
t585g07.x1 NCI-CGAP_GC6 Homo sapiens cDNA clone IMAGE:228108 3'
similar to SW:PILA_BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE
CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.
ACCESSION
A1632172
NID
94683502
VERSION
A1632172.1 GI:4683502
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 633)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
On May 7, 1998 this sequence version replaced gi:312411.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMNI at:
www.bio.lnlnl.gov/bdrp/image/image.html

FEATURES
source
Seq primer: -400P from Glibco
High quality sequence stop: 446.
Location/Qualifiers
1..633
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15q26.1"
/clone="IMAGE:228108"
/clone_lib="NCI-CGAP_GC6"
/tissue_type="Pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Plasmid DNA from the normalized library
NCI-CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR amplified cDNAs from a pool of 5,000 clones made
from the same library (cloneIDs 1257096-1258631,
1469064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo.

BASE COUNT 169 a 109 c 108 g 246 t 1 others
ORIGIN

Query Match 89.6%; Score 20.6; DB 49; Length 633;
 Best Local Similarity 78.3%; Pred. No. 7.1;
 Matches 18; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 AARATGATGGATCTTCACAC 23
 Db 391 AARATGATGGATCTTCACAC 369

RESULT 12
 LOCUS AI324274 453 bp mRNA EST 23-DEC-1998
 DEFINITION mg72c02.y1 Striatagene mouse melanoma (#937312) Mus musculus cDNA
 Clone IMAGE:584258 5' similar to SW:PI1A_MOUSE P42337
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
 mRNA sequence.

ACCESSION AI324274
 MID 94058703
 VERSION AI324274.1 GI:4058703
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 453)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B.,
 Thibisang, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 Unpublished (1996)
 On Jan 14, 1998 this sequence version replaced gi:1798536.

TITLE JOURNAL
 COMMENT Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LIND; contact the
 IMAGE Consortium (info@image.lind.gov) for further information.
 MGI:358906
 This read is a RESEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the
 correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 403.

FEATURES
 SOURCE Location/Qualifiers
 1..453
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="21"
 /clone="IMAGE:584258"
 /clone_idb="Striatagene mouse melanoma (#937312)"
 /tissue_type="melanoma"
 /dev_stage="M2 cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: Skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. From M2 cells; a highly metastatic derivative of
 the K-1735 (mouse) melanoma. Average insert size: 1.0 kb;
 Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGCGACGAG
 3' -3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'"

BASE COUNT 139 a 81 c 109 g 122 t 2 others
 ORIGIN

Query Match 87.0%; Score 20; DB 45; Length 453;
 Best Local Similarity 76.3%; Pred. No. 14;

Matches 18; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
 OY 1 AARATGATGGATCTTCACAC 23
 Db 389 AARATGATGGATCTTCACAC 411

RESULT 13
 LOCUS AA298585 343 bp mRNA EST 18-APR-1997
 DEFINITION EST114216 Hsc172 cells II Homo sapiens cDNA 5' end similar to
 phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
 sequence.

ACCESSION AA298585
 MID 91950938
 VERSION AA298585.1 GI:1950938
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 343)
 Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fullner, R.A.,
 Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
 White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, M.-C.,
 Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fline, L.D.,
 Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.,
 Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
 Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, P.T., Pellicciolo, S.M.,
 Moren, P., Palamara, R.F., McDonald, L.A., Nguyen, D.T., Pellicciolo, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
 Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dinko, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,
 He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.R.,
 Kozak, D.L., Kunsch, C., Hung, J., Li, H., Weissner, P.S., Olsen, H.,
 Raymond, L., Wel, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
 Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
 Fraser, C.M., and Venter, J.C.

TITLE JOURNAL
 COMMENT Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl), 3-174 (1995)
 96026280
 On Sep 12, 1996 this sequence version replaced gi:1392803.
 Other ESTs: THCA68479
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@tigr.org
 Information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.

FEATURES
 SOURCE Location/Qualifiers
 1..343
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):1179595"
 /db_xref="taxon:9606"
 /clone_idb="HSC172 cells II"
 /cell_type="fibroblast"
 /cell_line="HSC172 (60PDL)"
 /dev_stage="fetal"
 /note="Organ: Lung; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI"

BASE COUNT 132 a 61 c 64 g 84 t 2 others
 ORIGIN

Query Match 82.6%; Score 19; DB 31; Length 343;
 Best Local Similarity 73.9%; Pred. No. 39;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Cyt nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:13 ; Search time 539.84 Seconds
(without alignments)
206.192 Million cell updates/sec

Title: US-09-325-095-14

Perfect score: 35
Sequence: 1 GATGATGCGCCARCTGTTTCATATGATTTTGCCCA 35

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:*
1: gb_bal:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_p11:*
8: gb_p12:*
9: gb_prl:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_to:*
13: gb_st:*
14: gb_sy:*
15: gb_sy:*
16: gb_un:*
17: gb_v1:*
18: em_fun:*
19: em_hlg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_v1:*
34: gb_hlg1:*
35: gb_hlg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	38	5	A37242 Sequence 11

2	31.4	89.7	3452	4	AF001076	AF001076 Gallus ga
3	31.4	89.7	3389	17	AF001075	AF001075 Avian sar
4	29.8	85.1	3207	3	BOVPHOS3KN	M93252 Bovine phop
5	29.8	85.1	3498	5	A37232	A37232 Sequence 1
6	28.2	80.6	3424	10	HSPH13K	229090 H.sapiens m
7	28.2	80.6	3207	10	HSU79143	U79143 Human phosp
8	28.2	80.6	3207	12	MMU03279	U03279 Mus musculu
9	26.6	76.0	3868	10	HSP110DEL	Y10055 H.sapiens m
10	26.6	76.0	3525	10	HSU57843	U57843 Human phosp
11	26.6	76.0	5220	11	HSU86453	U86453 Human phosp
12	25	71.4	3132	12	MMU86587	U86587 Mus musculu
13	24.8	70.9	3213	10	S67334	S67334 phosphatidy
14	24.6	70.3	3712	36	DMPHOS3XT	Y09070 D.melanog
15	23.8	68.0	3213	12	RNO012462	AC002292 Genomic s
16	23	65.7	120787	8	AC002292	U97662 Chlamydomon
17	22.8	65.1	2200	8	CRU97662	Y11312 H.sapiens m
18	22.6	64.6	7654	10	HSC2P13K	Y13892 Homo sapien
19	22.6	64.6	5296	10	HSP13K1N	AC004199 Homo sapi
20	22.2	63.4	40622	11	AC004199	U53568 Homo sapien
21	22.2	63.4	37160	11	HSU53588	U23478 Dictyostell
22	22.2	63.4	4758	36	DDU23478	AC005543 Homo sapi
23	22.2	63.4	79810	42	AC005543	A17975 Glucose-6-p
24	21.8	62.3	1696	5	A17975	AC005896 Arabidops
25	21.8	62.3	119001	8	ATAC005896	AC005384 Homo sapi
26	21.8	62.3	178347	11	AC005384	AC006442 Homo sapi
27	21.4	61.1	155935	11	AC006442	AC003026 Human Chr
28	21.4	61.1	68355	11	HUAC03026	AC005636 Drosophil
29	21.4	61.1	70702	35	AC005636	AJ000008 Homo sapi
30	21.2	60.6	4880	9	HSC2P13K1	AB009636 Rattus no
31	21.2	60.6	5990	12	AB009636	AC007771 Drosophil
32	21.2	60.6	62980	35	AC007771	AC005555 Drosophil
33	21.2	60.6	101110	37	AC005555	D37799 Bacillus su
34	21.2	60.6	8054	1	BACAMOR000	Z09111 Bacillus su
35	21	60.0	208230	1	BSUB00008	AF012285 Bacillus
36	21	60.0	46864	2	AF012285	A47315 Sequence 6
37	21	60.0	662	5	A47315	X79944 Human papil
38	21	60.0	612	17	HOV5731	X32892 D.melanog
39	21	60.0	7215	36	DMP13K6BD	DB3984 Sulciculus di
40	21	60.0	15329	36	SMCIDOLM	U52192 Drosophila
41	21	60.0	6831	37	DMU52192	Z46973 H.sapiens m
42	20.8	59.4	2990	10	HSP1R1	M73514 Influenza A
43	20.8	59.4	2341	17	FLAHANP2	X74468 Human papil
44	20.8	59.4	7412	17	HPV15	AC006397 Homo sapi
45	20.8	59.4	91460	42	AC006397	

ALIGNMENTS

RESULT 1
A37242 38 bp DNA PAT 05-MAR-1997
LOCUS Sequence 11 from Patent WO9403609.
DEFINITION A37242
ACCESSION 92284353
VERSION A37242.1 GI:2294353
KEYWORDS
SOURCE
ORGANISM
unclassified.
unclassified.
unclassified.

REFERENCE 1 (bases 1 to 38)
AUTHORS Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUCHARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 11 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
Location/Qualifiers
1..38
source
/organism="unclassified"
/db_xref="taxon:32644"

BASE COUNT 7 a 6 c 8 g 11 t 6 others
ORIGIN

Query Match 94.38; Score 33; DB 5; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.00038;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 35
|||||
2 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 35

RESULT 2
AF001076 3452 bp mRNA VRT 08-JUL-1997
LOCUS
DEFINITION Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
complete cds.
ACCESSION AF001076
NID 92245505
VERSION AF001076.1 GI:2245505
KEYWORDS
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3452)
AUTHORS Chang,H.W., Aoki,M., and Vogt,P.K.
TITLE Tschilis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
JOURNAL Transformation of chicken cells by the gene encoding the catalytic
MEDLINE subunit of PI 3-kinase
REFERENCE Science 276 (5320), 1848-1850 (1997)
AUTHORS 2 (bases 1 to 3452)
TITLE Chang,H.W., Aoki,M. and Vogt,P.K.
JOURNAL Direct Submission
MEDLINE Submitted (23-APR-1997) Molecular & Experimental Medicine, The
AUTHORS Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
JOURNAL BCC-239, La Jolla, CA 92037, USA
LOCATION/Qualifiers

FEATURES
source 1..3452
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="brain"
1..150
151..3354
/gene="c-p3k"
/note="p3k proto-oncogene"
151..3354
/note="c-p3k"
/codon_start=1
/product="phosphoinositide 3-kinase catalytic subunit"
/protein_id="AAB62534.1"
/db_xref="PDB:92245506"
/translation="MPRSSGELMGILHMPRIIVLECLLPNGMIVTLLECLREATLT
IKHELEFKARKYPLVQLDDESSYIEFVQEARREFEETRLCDLRFPLVY
EPKNGREKILNREIGFALGMPICERFDVPEQVRNLLVCKEAVDLRANPH
SRLLVCPVWESPPELPRIVNKLKGOIIVYIVYISNNDKQYTLKINDCPE
QVAAIRKRTSRMLSSSEDLKCVLEIQKYLKVGCGCEYLLEKTPLSQKVIKNSC
IMGRPNMLMAKESLYTQPLDTFTMPYSRISTATPYNGEATASLWINSAL
RILLCATYVNNVNIRODKIYVGTGTYHGEPLCDVNTQVPCSNPRNEMTSYMY
IPDLPRARLCLTICSVKGRKAEEHCPLAMGNINMFDTDTLVGKALNLAAYH
GLEDLNPITGVGNSNPKEPCLEEDFDMFSNPKPPDMVIEHANMTISRELGNY
SVAGLSNRIRARDELRESDEQARICTRBPISLSEITQENDPFRSHRHCYVNPILP
KLLSVKWSRDEVAQMTCLVQMPPIKPEQAMELDCNPDVVAFAVRCLEKLT
DDLSQYLQIVLVLYEDLYQVLFLLKALTNQRIQGHFFMHLKSMHKKTVSQ
RGLLEISYCRACGMVTLKLSRQVEAMEKILNTDLKQKDEIQKVMKFLVEQNR
RPFMDALQGISPLNPAHOLGNLRLECRINSASAKPLMLNENPDIIMSELFOHNE
IIFKNGDDLRODMTLQIIRIMENIMONGIDLRLM.PYGLSTIGDGVGLIEVRSST
IMQIOCKGKLGALQFNSTHLOWLMDKNGEYDAIDLFTSCAGCYATITLISIG
DRNSNIWAKDQGLHIDGHLDRKKKFGYKREKRVPLTQDLVYISKAOECT
KREFEFQEMCKRAYLAIHQHNLINFLSMVLGSGMPELQSFDDIIVIRKTLADLK
TEQEALEYFMKONNDAHGQWTKMDIIEFTIIOHNLN"
3358..3452

3'UTR

BASE COUNT 1073 a 672 c 783 g 924 t
ORIGIN

Query Match 89.7%; Score 31.4; DB 4; Length 3452;
Best Local Similarity 82.9%; Pred. No. 0.003;
Matches 29; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

1 1 GATGATGGCCACTGTTTCATATGATTTGGCCA 35
|||||
2 2923 GATGATGGCCACTGTTTCATATGATTTGGCCA 2957

RESULT 3
AF001075 3389 bp mRNA VRL 08-JUL-1997
LOCUS
DEFINITION Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
partial cds.
ACCESSION AF001075
NID 92245502
VERSION AF001075.1 GI:2245502
KEYWORDS
SOURCE Avian sarcoma virus 16.
ORGANISM Avian sarcoma virus 16.
VIRUSES: Retroid viruses; Retroviridae; Avian type C retroviruses.
REFERENCE 1 (bases 1 to 3389)
AUTHORS Chang,H.W., Aoki,M., Fruman,D., Auger,K.R., Bellacosa,A.,
Tschilis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
JOURNAL subunit of PI 3-kinase
MEDLINE Science 276 (5320), 1848-1850 (1997)
AUTHORS 2 (bases 1 to 3389)
TITLE Chang,H.W., Aoki,M. and Vogt,P.K.
JOURNAL Direct Submission
MEDLINE Submitted (23-APR-1997) Molecular & Experimental Medicine, The
AUTHORS Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
JOURNAL BCC-239, La Jolla, CA 92037, USA
LOCATION/Qualifiers

FEATURES
source 1..3389
/organism="Avian sarcoma virus 16"
/db_xref="taxon:60629"
<1..3252
/gene="gag-v-p3k"
/note="fusion gene of gag and v-p3k"
1..78
/gene="gag-v-p3k"
/note="derived from gag gene"
<1..3252
/gene="gag-v-p3k"
/codon_start=1
/product="gag-v-phosphoinositide 3-kinase catalytic
subunit fusion protein"
/protein_id="AAB62532.1"
/db_xref="PDB:92245503"
/translation="QQLRTAPSTLTTPGEIITKYVDROKASHLMPRIIVLECLLPNG
MIVTLLECLREATLTIKHELEFKARKYPLVQLDDESSYIEFVQEARREFEETRLCDLRFPLVY
EPKNGREKILNREIGFALGMPICERFDVPEQVRNLLVCKEAVDLRANPH
SRLLVCPVWESPPELPRIVNKLKGOIIVYIVYISNNDKQYTLKINDCPE
QVAAIRKRTSRMLSSSEDLKCVLEIQKYLKVGCGCEYLLEKTPLSQKVIKNSC
IMGRPNMLMAKESLYTQPLDTFTMPYSRISTATPYNGEATASLWINSAL
RILLCATYVNNVNIRODKIYVGTGTYHGEPLCDVNTQVPCSNPRNEMTSYMY
IPDLPRARLCLTICSVKGRKAEEHCPLAMGNINMFDTDTLVGKALNLAAYH
GLEDLNPITGVGNSNPKEPCLEEDFDMFSNPKPPDMVIEHANMTISRELGNY
SVAGLSNRIRARDELRESDEQARICTRBPISLSEITQENDPFRSHRHCYVNPILP
KLLSVKWSRDEVAQMTCLVQMPPIKPEQAMELDCNPDVVAFAVRCLEKLT
DDLSQYLQIVLVLYEDLYQVLFLLKALTNQRIQGHFFMHLKSMHKKTVSQ
RGLLEISYCRACGMVTLKLSRQVEAMEKILNTDLKQKDEIQKVMKFLVEQNR
RPFMDALQGISPLNPAHOLGNLRLECRINSASAKPLMLNENPDIIMSELFOHNE
IIFKNGDDLRODMTLQIIRIMENIMONGIDLRLM.PYGLSTIGDGVGLIEVRSST
IMQIOCKGKLGALQFNSTHLOWLMDKNGEYDAIDLFTSCAGCYATITLISIG
DRNSNIWAKDQGLHIDGHLDRKKKFGYKREKRVPLTQDLVYISKAOECT
KREFEFQEMCKRAYLAIHQHNLINFLSMVLGSGMPELQSFDDIIVIRKTLADLK
TEQEALEYFMKONNDAHGQWTKMDIIEFTIIOHNLN"

3'UTR

[illegible]

BASE COUNT	ORIGIN	1028 a	581 c	680 g	918 t
Query Match		85.18;	Score 29.8;	DB 3;	Length 3207;
Best Local Similarity		80.0%;	Pred. No. 0.014;		
Matches	28;	Conservative	5;	Mismatches	2;
				Gaps	0;
				Indels	0;
QY	1	GATGATGGCCARCTGTTTCATATGATTTGGCCA	35		
Db	2773	GATGATGGACACAGTGTTCATATGATTTGGACA	2807		
RESULT	5				
LOCUS	A37232	3498 bp	DNA	PAT	05-MAR-1997
DEFINITION	Sequence 1 from Patent WO9403609.				
ACCESSION	A37232				
NID	92294345				
VERSION	A37232.1	GI:2294345			
KEYWORDS	unidentified.				
SOURCE	unidentified.				
ORGANISM	unclassified.				
REFERENCE	1 (bases 1 to 3498)				
AUTHORS	Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.				
TITLE	ECARBOXYLIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE OR PROTEIN KINASE AND ASSAYS USING THEM				
JOURNAL	Patent: WO 9403609-A 1 17-FEB-1994;				
COMMENT	IMP CANCER RES TECH (GB)				
FEATURES	Other publication JP 8503124T 960409.				
Source	Location/Qualifiers				
	1..3498				
	/organism="unidentified"				
	/db_xref="taxon:32644"				
	1..3207				
	/note="unnamed protein product"				
	/cdon_start=1				
	/protein_id="CAA02284.1"				
	/db_xref="PID:e304808"				
	/db_xref="PID:g2294346"				
	/db_xref="GI:2294346"				
	/translation="MPRRPSGELMGILHMPRLIVBCLPNCIVTLDECREATLIT IKHELEAREKPYLHOLDDESSITIVSVTQEAEREFEFTETRLCDRLPQPLKVT EPYGNREKILNELIEGFIAGMPCEDDMVPEVDERRNLINVCAPVLDRLNDSPE SRMAYVYPNVVESPELPHKHIYKMLDKGQIIVIVVSVSNPKOKYTLINHCYBE QVIAEARIKTRKRSNLSSSEQLKCVLEQSYKILKYGCEYFLKPLSOYKIRSC IMGEMPRMLMLAKESLIYSOLPMDCEFMPSYRSIRATYNGEFTKSLAYINSL RILICATYVNNVIRIDRIYTVGTIGHGEPICLDVNTQRYVCSNRRNEMVNDIY IPLDPLRIARCLSTISVYGRKGAKEECPILAMNINLEFDITLVISGKMLNMPVH GLDPLRIADIVGTSNPKEETPCLEDEDMVSPVKEPDMVIEEHANVSAREGSESY SHAGLSNRLARNDLENDEKQELALICTRPLSEITQEKDFLMSHRHYCVTIPLEILP KLLSYKVRNLRDEVAQWYCLVMDMPKPKPOAEMALDCNPPDPRVAFVARGLEKYL DKLTSOYLQIOLVYKREOYLQNLVYALFKLKTNRQKKEFFGMHLSKMHKNTVSO RRLTLESYCRACGMVTKHLNRQVEAMEKILNLTDLQKQKDETVQYOKPLVEQNR RPFMDLQGFSLPPLNPAHQUNLRRECKIMSSAKRPLNLNENPDIKSELHQUNNE IIRKNGDGLQMDLTLQIRIMENINQOGLDJRLMPYGLSLGDCGLIEVYNSHT IMOIOCGGLKGAQGFNSHTLHOMLNKNGEITYDAIDIFTSAGCYVATFLLGIG DRNNSINMWKDDQLFIIDFGHFLDHKKRKFYGRREVPVLQDFDIATVSKAOECT KTEDEFERQCMKAYALATROHANLPFLNLSMAGSGMPELQSFDDIATIRKTLADLK TEODALEYFKQKMDAHHGWTTRKMDIIFITIKOHALN"				
BASE COUNT		1147 a	626 c	722 g	1003 t

FEATURES	source	Location/Qualifiers
JOURNAL	J. Biol. Chem. 272 (31), 19236-19241 (1997)	
MEDLINE	97382246	
REFERENCE	2 (bases 1 to 3132)	
AUTHORS	Vojtek, A. B. and Cooper, J.	
TITLE	Direct Submission	
JOURNAL	Submitted (22-JAN-1997) Biological Chemistry, University of Michigan, 3323 MSRB III, Box 0636, Ann Arbor, MI 48109-0636, USA	
FEATURES		
source		
1. .3132	/organism="Mus musculus"	
/db_xref="taxon:10090"		
/tissue_type="spleen"		
1. .3132		
/codon_start=1		
/product="phosphatidylinositol 3-kinase catalytic subunit p110 delta"		
/protein_id="AAC25676.1"		
/db_xref="PID:g2331238"		
/db_xref="GI:2331238"		
/translation="MPPEVDDPMEFWTKESQSVVDFLLPTGVLYNFPVSRNMLSTIKQYLHRAQOEPLPFHMSDPEAYVFCVNTAQOELDEDRCLDQPLPYRLALVLYPILALVYKALNDSILILAKGLHEPDSLRPEVNDFTKKRQCEAAARQOLGWAEMLQSPLELPESAPGRAGLLRVSNRALLVYKFNPTKSESPFPQVSTKMPALAMWCALEKRAKTVROPDYEOPEERIALDVNGRHETLYGNPYLCHFOYICSCISHSGLTHLTRVHSSILAMRDEOSNPAPQVQKPRAKPPPIYAKRPSSVLSLQPESTILEGRKYNVADESMKLVQAGLPHGNEMLCRTVSSSVNCSBPYMKQKLEFPISYDDLPRMARLCEALYVVERAKKARSTKRSKADCPIDAMANMLDQDLQTLGERCLYMPVSDEKELFLNAPAGVRRGNPTESAAALVILYIEVAPHPVEPALEKILLEGRRGRTIEEELQRLPIERRSGELYEHRKDLVYKMRHVPQHPPEALARLLYTKMKHEDVYADMTYLICSPPELPVLSALTELLBSPFDCYGSFALSLKRLTDELFOYLLVQVLTKESTLDCETKFLGLALANRRIGHFLFNLHSEMPVVALRFGILMEAYCGSTHNKLYVLMKOGALSKRLKALNDFYVSSQKTKRPQTKEMHMCROEYMEALSHLOSPLDSTLILEEVEQCEQTEFMSCKRPLMYISEEGASAGNVGILIFKNGDGLNODMTLLQIOMDVLMKREGDLRMTPYGCLPTGDRGTILEVILSHDILANTOLKMSMAAANAANKRALIKWKKNGEALDRAIEBFTLSCAGYCAVTVLYIGDRDSNMILRESQFLRIDPGEHNLGNKTKFGLINRRVPTLVLYDFVHAVIQOGTNNSEKFERRGYCERYVTLIRRHGDLFLHLFALMARAGDELSCSDIOYLIKDSLALGKTEEBALKHFRVKAFNALRESWKTKVNWLANHNSKDRQ"		
BASE COUNT	690 a 920 c 901 g 621 t	
ORIGIN		
Query Match	71.4%; Score 25; DB 12; Length 3132;	
Best Local Similarity	71.4%; Pred. No. 1.5;	
Matches	25; Conservative 5; Mismatches 5; Indels 0; Gaps 0.	
Db	2704 GAGACTGGGACGCTTCACATGATTTGGCCA 2738	
RESULT	13	
LOCUS	S67334	
DEFINITION	phosphatidylinositol 3-kinase p110 beta isoform-110 kda catalytic subunit [human, mRNA Partial, 3213 nt].	
ACCESSION	S67334	
NID	9455759	
VERSION	S67334.1	
KEYWORDS	GI:455759	

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 3213)
TITLE Hu.P., Mondino, A., Skolnik, E. Y. and Schlessinger, J.
 Cloning of a novel, ubiquitously expressed human phosphatidylinositol 3-kinase and identification of its binding site on p85
JOURNAL Mol. Cell. Biol. 13 (12), 7677-7688 (1993)
MEDLINE 94067128
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gidsbg 140879] from the original journal article. This sequence comes from Fig. 1.
FEATURES location/Qualifiers
source 1..3213
organism "Homo sapiens"
gene /db_xref="taxon:9606"
 1..3213
gene /gene="phosphatidylinositol 3-kinase p110 beta isoform, PI 3-kinase p110 beta"
 1..3213
CDS /gene="phosphatidylinositol 3-kinase p110 beta isoform, PI 3-kinase p110 beta"
 /note="110 kda catalytic subunit; This sequence comes from Fig. 1; PI 3-kinase p110 beta"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase p110 beta isoform"
 /protein_id="AA829081.1"
 /db_xref="PID:g455760"
 /db_xref="GI:453760"
 /translation="MCSFTMPAMADIDIMAVDSQIASDGSIPVDFLLPTGIYQL
 EYREAVATISYIKOMLVKNVPMFLNMDIDSMFCVNOIAYELEDDETRLCQV
 RPEPLVPLKTRSCDPEKGLDSKIGVIGLGHPEFDLNDPEVNEFRKKRFSSEKI
 LSLVGLSMDMLKOTYPRPEHPSIPREPLEDKELGKILVAVHEENCODVRSFOVSPMM
 NPIKVNELAQKRLTIHGKDEVSPIYVYQVSGRYVYVGDHPLIDFOYIRNCYARR
 ALPFIIVECCIKTKMTQEMLIAIEALINSSMLPLPLPKTRITISHYENNPPQ
 ILVAVGNKLTLEEVKAVHRAVGLFHGELLCTIVSEVSGKNDHINBELPDINIC
 DLPMARICFAVYAVLDKRVKSTKSTINSKQTIIRKAKVHYPVAVMTWDFEDNGC
 QLRGDIILHSSMSFPELEEMLMPTGOTINPTENATLHVAFENRKOPYYPPE
 DKITERKAETIASNSVAVSSRGKKPLPVKELIDRPLDLOECENMDLIWTLRODCR
 EITPQSPKRLILSKMKKEDVNAOLALDITWRLPPREALBELDFNYPPOYREYAV
 GICRQMSDELSQTLQVQVLYKEPLDCLSFLELRLGNRRIGQFLPHURSEV
 HPAVSVQFVILEAYCRGVSVGHMKVLSKQVEALNKLTKLNSLILKLAVALNRKGEK
 AMHTGLKOSAYRALSDLOSPLNCPVILSELVEYCKYDSKKPLVYNKKVEGSD
 SVAGTICRGGDLDRODMLTLMRLMDLMEAGLDMLPYGCLATGDRGLIEVST
 SERTADIGLSSNNSVAKTAAPFNKDALINKEEYSGDIDRLEPFTISGAGCYAVSYV
 LGIDRRSDINIVKRTIQGLFHIDFGHLITGFKSFGVIGKREVRVPIITDYDTHVQSK
 TGNTEKGRFRQCCEDAYILIRRGNTLFLPALMTAGLEPLTIVANDIYLDVSLAL
 GKSEBELKQFKQFDEALRESVTTKYNMAAHVTRKDYRS"
BASE COUNT 979 a 612 c 704 g 918 t
ORIGIN
 Query Match 70.9%; Score 24.8; DB 10; Length 3213;
 Best Local Similarity 76.7%; Pred. No. 1.9;
 Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Oy 6 TGGCAGCTGTTTCATATGAGATTTCGCCA 35
 |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
 Db 2790 TGGCGAGCTCTTCACATGACTTTGGACA 2819
RESULT 14
DMPHOS3KI DMPHOS3KI 3712 bp mRNA INV 07-JAN-1997
LOCUS D.melanogaster mRNA for phosphoinositide 3-kinase, p110.
DEFINITION Y09070
ACCESSION 91707447
NID Y09070.1 GI:1707447
VERSION phosphoinositide 3-kinase.
KEYWORDS fruit fly.
SOURCE Drosophila melanogaster

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:29 ; Search time 255.05 Seconds

(without alignments)
34.333 Million cell updates/sec

Title: US-09-325-095-14

Perfect score: 35
Sequence: 1 GATGATGCCARCTGTCATCATGATTTGGCCA 35

Scoring table: IDENTITY_MUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	94.3	38	1	Q59013
2	33	94.3	38	1	Q57019
3	29.8	85.1	3207	1	Q51155
4	29.8	85.1	3498	1	Q57012
5	28.2	80.6	3412	1	Q51156
6	26.6	76.0	3387	1	V16533
7	26.6	76.0	5220	1	V13340
8	26.6	76.0	5220	1	X15932
9	21.8	62.3	1696	1	Q20959
10	21.8	62.3	1455	1	T30577
11	21.4	61.1	3252	1	X24400
12	21	60.0	381	1	Q57522
13	21	60.0	662	1	T03504
14	21	60.0	3202	1	T47520
15	21	60.0	6831	1	T80200
16	21	60.0	3316	1	V44859
17	20.8	59.4	2121	1	V11860
18	20.6	58.9	1736	1	T37149
19	20.2	57.7	1461	1	T30576
20	19.8	56.6	6693	1	V06570
21	19.6	56.0	934	1	T05843
22	19.6	56.0	5285	1	T80199
23	19.6	56.0	5061	1	V42920
24	19.6	56.0	4956	1	X13228
25	19.6	56.0	1500	1	X20016
26	19.6	56.0	1399	1	X20017
27	19.2	54.9	674	1	T03503
28	19.2	54.9	665	1	T03505
29	18.2	54.9	3602	1	V01877
30	19.2	54.9	2487	1	V01878
31	19.2	54.9	3324	1	V01879
32	19.2	54.9	2451	1	V01876
33	19.2	54.9	410	1	V15596
34	19.2	54.9	12565	1	V62392
35	19.2	54.3	674	1	T03506
36	19	54.3	1193	1	T79975
37	19	54.3	1958	1	V10328
38	19	54.3	1690	1	V10327
39	19	54.3	1714	1	V19790
40	19	54.3	2200	1	V44929
41	19	54.3	2200	1	V28288
42	19	54.3	2200	1	V32648
43	19	54.3	110000	1	X20248_03

ALIGNMENTS

44 19 54.3 5940 1 X23526 O. longistaminta X
45 19 54.3 8416 1 X23523 O. longistaminta

```
RESULT 1
ID 059013 standard; DNA: 38 BP.
AC 059013:
DE 12-APR-1994 (first entry)
KW Primer for p110 cDNA.
KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ss: amplification.
OS Synthetic.
PN W09321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1992; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Dhond R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker PJ, Volinia S, Waterfield MD;
PI WPI: 93-351738/44.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Example 1: Page 38; 146pp: English.
CC An SGBAF-1 cell line was established by transfection of bovine
CC adrenal cortex zona fasciculata cells with pSVneo. Total RNA was
CC isolated from the SGBAF-1 and a cDNA library constructed. The
CC unamplified library was plated on E. coli K12 PLF-F' and screened
CC with the labelled primer shown and the primer of Q59012. Hybridising
CC clones were sequenced.
CC See also Q51155-6, Q59012-23 and Q57522-3.
SQ Sequence 38 BP: 7 A; 6 C; 8 G; 11 T;

Query Match 94.3%; Score 33; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 2.8e-05;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GATGATGCCARCTGTCATCATGATTTGGCCA 35
DB 1 GATGATGCCARCTGTCATCATGATTTGGCCA 35

RESULT 2
ID 057019 standard; DNA: 38 BP.
AC 057019:
DE 31-AUG-1994 (first entry)
DE P14ns 3-kinase probe #2.
KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW P14ns 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.
OS Synthetic.
PN W09403609-A.
PD 17-FEB-1994.
PF 05-AUG-1993; G01651.
PR 05-AUG-1992; GB-016654.
PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI WPI: 94-065687/08.
PT Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in
PT cell growth regulation and for treating cancers
PS Example 1: Page 14; 71pp: English.
CC The sequences given in Q57018-19 are probes which were used in the
CC isolation of fusion sequences comprising the phosphatidylinositol
CC (P14ns) 3-kinase under the regulatory control of the nmt promoter. ..
```

CC These sequences were transformed into Schizosaccharomyces pombe cells
 CC in an embodiment of the invention. In the presence of thiamine the
 CC promoter is inactive and the cells carrying the PKC plasmids grow as
 CC the parental strain. In the absence of thiamine the nmt promoter
 CC functions and the PKC is induced. PKC activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 94.3%; Score 33; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 2.8e-05;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 GATGATGCCACCTGTTTCATATGATTTGGCCA 35
 1 GATGATGCCACCTGTTTCATATGATTTGGCCA 35

RESULT 3
 ID 051155 standard; cDNA; 3207 BP.
 AC 051155;
 DE 12-APR-1994 (first entry)
 UI P110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antaagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB: R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K61a. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (ant)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 SO Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 85.1%; Score 29.8; DB 1; Length 3207;
 Best Local Similarity 80.0%; Pred. No. 0.0018;
 Matches 28; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GATGATGCCACCTGTTTCATATGATTTGGCCA 35
 2773 GATGATGCCACCTGTTTCATATGATTTGGCCA 2807

RESULT 4
 ID 057012 standard; cDNA to mRNA; 3498 BP.
 AC 057012;
 DE 31-AUG-1994 (first entry)
 UI Ptdins 3-kinase 110 KD catalytic subunit cDNA.
 KW 110 KD catalytic subunit; phosphatidy1 inositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /product= p110

WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB: R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 KD catalytic subunit of the phosphatidy1
 CC inositol (ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 85.1%; Score 29.8; DB 1; Length 3498;
 Best Local Similarity 80.0%; Pred. No. 0.0019;
 Matches 28; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

1 GATGATGCCACCTGTTTCATATGATTTGGCCA 35
 2773 GATGATGCCACCTGTTTCATATGATTTGGCCA 2807

RESULT 5
 ID 051156 standard; cDNA; 3412 BP.
 AC 051156;
 DE 12-APR-1994 (first entry)
 UI Human p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antaagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss.
 OS Human.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.


```

FT primer_bind /tag= a complement (34, .58)
FT primer_bind /tag= b /note= "Primer for HDGF-2 recombinant prodn."
PD WO9639485-A1.
PN 12-DEC-1996.
PF 05-JUN-1995; U06731.
PR 05-JUN-1995; WO-U06731.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Kunsch CA, Rosen CA;
DR MPI; 97-043108/04.
DR P-PSMB; W09404.
PT DNA encoding human-derived growth factor polypeptide - useful to
PT promote wound healing as result of burns or ulcers
PS Claim 1, Fig 1A-C; 54pp; English.
CC A cDNA clone (747520) codes for human hepatoma-derived growth
CC factor-2 (HDGF-2) (W09404), a protein that can be used to
CC stimulate tissue repair and tissue growth. It was discovered in
CC a human umbilical vein endothelial tissue cDNA library, and may
CC also be obtained from heart, brain and skeletal muscle. HDGF-2
CC polynucleotides can be incorporated into vectors for prodn. of
CC recombinant HDGF-2 in host (e.g., E. coli, S19) cells. They can
CC also be used for the gene therapy of diseases related to HDGF-2
CC underexpression and for the design of diagnostic probes.
SQ Sequence 3202 BP; 817 A; 816 C; 819 G; 747 T;

Query Match 60.0%; Score 21; DB 1; Length 3202;
Best Local Similarity 68.6%; Pred. No. 8;
Matches 24; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 GATGATGCCARCTGTTTCATATGATGATTTGGCA 35
Ub 937 GATCTGGCCACTGTCCTGATGAGATGTCCA 903

RESULT 15
T80200
TD T80200 standard; cDNA: 6831 BP.
AC T80200;
UT 22-JUN-1998 (first entry)
DE Phosphatidyl inositol 3-kinase cdk cDNA.
KW Phosphatidyl inositol 3-kinase; signal transduction; cell cycle;
KW autophagy; inflammatory joint disease; cell proliferation; cancer;
KW psoriasis; restenosis; atherosclerosis; therapy; diagnosis; cpk; ss.
XX Drosophila melanogaster.
FH Key Location/Qualifiers
FT CDS 148..5778
FT /tag= a
PN WO9731650-A1.
PD 04-SEP-1997.
PF 12-FEB-1997; U02193.
PR 29-FEB-1996; US-609049.
PA (REGC ) UNIV CALIFORNIA.
PI Chen Y, Moiz L, Williams LT;
PI WPI; 97-448442/41.
PR P-PSDB; W38757.
PT New isolated phosphatidyl inositol-3 kinase polypeptide - used to
PT develop products for diagnosis and therapy, particularly for
PT proliferative disorders, e.g. inflammatory joint diseases, or cancer
PT Example 1; Fig 9; 77pp; English.
CC This cDNA sequence codes for cpk (see W38757), a Drosophila
CC polypeptide that belongs to a novel class of phosphatidyl inositol
CC 3-kinases that contain a C2 domain, are capable of phosphorylating
CC a D3 hydroxyl of an inositol ring in phosphatidyl inositol and
CC phosphatidyl inositol 4-phosphate, but not in phosphatidyl inositol
CC 4,5-bisphosphate, and which are involved in cell signalling cascades
CC that control e.g. cell cycle progression and intracellular protein
CC sorting. Short fragments of cpk cDNA were obtained from a
CC Drosophila cDNA library by PCR (see T80195-96). These short
CC fragments were used to screen the cDNA library to obtain larger
CC fragments, and missing 5' ends were obtained by RACE. A
CC recombinant host cell, transfected with a vector comprising a cpk

```

```

CC nucleic acid is claimed. Novel phosphatidyl inositol 3-kinases
CC can be used to screen for agonists/antagonists of activity and
CC in a claimed method of treating a disorder caused by dysregulation
CC of a growth factor activation signalling cascade. Antagonists
CC may reduce Ras activation allowing treatment of proliferative
CC disorders such as atherosclerosis, inflammatory joint disease,
CC psoriasis, restenosis following angioplasty, and cancer.
SQ Sequence 6831 BP; 1829 A; 1685 C; 1669 G; 1648 T;

Query Match 60.0%; Score 21; DB 1; Length 6831;
Best Local Similarity 72.4%; Pred. No. 11;
Matches 21; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 7 GGCCARCTGTTTCATATGATGATTTGGCCA 35
DB 4498 GGTCACCTGTTTCACATGACCTTGGCAA 4526

Search completed: September 13, 1999, 15:59:29
Job time: 4796 sec

```

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Search time: 15:45:55 : Search time 1694.61 Seconds
(Without alignments)
40.740 Million cell updates/sec

US-09-325-095-14

Title: 35
Perfect score: 1
Sequence: 1 GATGATGCCARCTGTCATCATMGATTTGGCCA 35

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:
1: em_est1:
2: em_est2:
3: em_est3:
4: em_est4:
5: em_est5:
6: em_est6:
7: em_est7:
8: em_est8:
9: em_est9:
10: em_est10:
11: em_est11:
12: em_est12:
13: em_est13:
14: em_est14:
15: em_est15:
16: em_est16:
17: em_est17:
18: em_est18:
19: em_est19:
20: em_est20:
21: em_est21:
22: em_est22:
23: em_est23:
24: em_est24:
25: em_est25:
26: em_est26:
27: em_est27:
28: em_est28:
29: em_est29:
30: em_est30:
31: em_est31:
32: em_est32:
33: em_est33:
34: em_est34:
35: em_est35:
36: em_est36:
37: em_est37:
38: em_est38:
39: em_est39:
40: em_est40:
41: em_est41:
42: em_est42:
43: em_est43:
44: em_est44:
45: em_est45:
46: em_est46:
47: em_est47:
48: em_est48:
49: em_est49:
50: em_est50:
51: em_est51:
52: em_est52:
53: em_est53:

54: em_est22:
55: em_est23:
56: em_est24:
57: em_est25:
58: em_est26:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	28.2	80.6	453	45	A1324274	A1324274 mg72c02.y
2	25	71.4	476	29	A1174738	A1174738 mt12e11.r
3	24.8	70.9	328	21	R12465	R12465 yf56h07.r1
4	24.8	70.9	391	29	A1190747	A1190747 zp8h405.r
5	22.4	64.0	479	25	N76563	N76563 yv40a03.r1
6	21.6	61.7	447	42	A1155136	A1155136 ud59f06.r
7	21.4	61.1	378	29	A1193439	A1193439 zr41b09.s
8	21.2	60.6	263	21	R18590	R18590 yf96b10.r1
9	21.2	60.6	509	21	R21085	R21085 y952f07.r1
10	21.2	60.6	330	22	R22699	R22699 y908e01.r1
11	21.2	60.6	574	37	AB009107	AB009107 AB009107
12	21	60.0	336	20	Z25224	Z25224 HSBABE042.S
13	20.8	59.4	497	29	A1172940	A1172940 mt01g12.r
14	20.8	59.4	446	29	A1175018	A1175018 ms82a03.r
15	20.8	59.4	517	29	A1175929	A1175929 ms92d03.r
16	20.8	59.4	494	30	AA243786	AA243786 zr67f12.r
17	20.8	59.4	452	31	AA287448	AA287448 z552d04.r
18	20.8	59.4	393	39	AA835841	AA835841 oc76d05.s
19	20.8	59.4	443	41	A1041241	A1041241 ub04d07.r
20	20.8	59.4	461	41	A1047071	A1047071 ub06h10.r
21	20.8	59.4	464	42	A1157114	A1157114 ud16d01.r
22	20.8	59.4	402	43	A1197559	A1197559 ud99a10.r
23	20.6	58.9	509	35	AA596554	AA596554 vms7d10.r
24	20.6	58.9	624	41	A1054648	A1054648 coa0001J
25	20.6	58.9	412	46	A1448463	A1448463 mg45d01.x
26	20.6	58.9	359	47	A1465317	A1465317 mu66h05.x
27	20.6	58.9	394	48	A1608079	A1608079 va08c12.y
28	20.6	58.9	328	48	A1661235	A1661235 va08c12.x
29	20.6	58.9	341	49	AV023366	AV023366 AV023366
30	20.4	58.3	338	21	T86707	T86707 yd86f07.s1
31	20.4	58.3	551	42	A1108767	A1108767 GH08006.5
32	20.4	58.3	509	42	A1108775	A1108775 GH08018.5
33	20.2	57.7	346	20	D24320	D24320 RICR1727A.R
34	20.2	57.7	373	22	H02035	H02035 yj38g08.s1
35	20.2	57.7	221	39	AA876400	AA876400 nw77h04.s
36	20.2	57.7	412	42	A1091644	A1091644 oc01g10.x
37	20.2	57.7	333	50	AV046640	AV046640 AV046640
38	20	57.1	329	48	A1600497	A1600497 486075F12
39	20	57.1	693	50	AU060839	AU060839 AU060839
40	20	57.1	591	51	A1734618	A1734618 60603A09
41	19.8	56.6	419	22	R26169	R26169 yb39g11.s1
42	19.8	56.6	352	40	AA933693	AA933693 oc074d02.s
43	19.6	56.0	139	29	AA168605	AA168605 nc30h01.r
44	19.6	56.0	624	29	AA901889	AA901889 NCP2EGT7
45	19.6	56.0	596	39	AA901891	AA901891 NCM4F4T3

ALIGNMENTS

RESULT 1
LOCUS A1324274 453 bp mRNA
DEFINITION mg72c02.y1 Striatogene mouse melanoma (#937312) Mus musculus cDNA
clone IMAGE:584258 5' similar to SW:P1A_MOUSE P42337
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.
ACCESSION A1324274

REFERENCE	AUTHORS		Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia: Eutheria: Rodentia: Sciurognath: Muridae: Murinae: Mus.
	1 (bases 1 to 476)		Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gessel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Thelning,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	On Apr 14, 1993 this sequence version replaced gi:693229.		
FEATURES	source		Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouse@watson.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.lnl.gov) for further information. MGI:381700 Seq primer: -28M13 rev2 from Amersham High quality sequence stop: 475. Location/Qualifiers 1..476 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /map="5 p15.2-p15.1" /clone="IMAGE:620876" /clone_id="Soares mouse 3bMS" /sex="male" /tissue_type="Spleen" /dev_stage="4 weeks" /lab_host="DH10B" /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCGCTGTGTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by Bento Soares and M.Patima Bonaldo."
BASE COUNT	117 a 132 c 127 g 99 t		1 others
ORIGIN	Query Match		71.4%; Score 25; DB 29; Length 476;
	Best Local Similarity		71.4%; Pred. No. 1.4;
	Matches 25; Conservative		5; Mismatches 5; Indels 0; Gaps 0;
QY	1	GATGATGGCCARCTGTTTCAYATGCAATTTGGCCA 35	
	11	: : : : :	
Db	37	GAGACTGGCAGCGCTTCCACATGTGATTTGGCCA 71	
RESULT	3		
LOCUS	R12466		328 bp mRNA EST 12-APR-1995
DEFINITION	Y556h07.r1 Soares infant brain INIB Homo sapiens cDNA clone IMAGE:26444.5' similar to SP:P100_BOVIN P32871 PHOSPHATIDYLINOSITOL 3'-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.		
ACCESSION	R12466		
NID	9765542		
VERSION	R12466.1		GI:765542
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eukaryota; Primates; Catarrhini; Homiidae; Homo.		

REFERENCE
AUTHORS
1 (bases 1 to 328)
Hillier,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.
JOURNAL COMMENT
The Mashu-Merck EST Project unpublished (1995)

Contact: Wilson RK
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2060
High quality sequence stops: 256 Source: IMAGE Consortium, LNL This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2060 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 256.

FEATURES
source Location/Qualifiers
..328
/organism="Homo sapiens"
/db_xref="GDB:398791"
/db_xref="taxon:9606"
/clone="IMAGE:26444"
/clone_1lb="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lambda BA; Site_1: Not I; Site_2: Hind III; 1st strand cDNA was primed with a Nobo I - oligo(dT) primer [5'; AACGGAGAAATTCGCCGCACAGATTCTTTTTTTTTT 3']; double-stranded cDNA was ligated to Hind III adaptors (Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the Lambda BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

DNA COUNT
ORIGIN 86 a 61 c 83 g 98 t

Query Match
Best Local Similarity 70.9%; Score 24.8; DB 21; length 328;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 6 TGGCCARCGTGTCAYATWGAATTTGCCCA 35
||||| ||||| ||||| ||||| ||||| ||
Db 110 TGGCAGCGCTCTCCACACTGACTGTGGACA 139

RESULT 4
LOCUS AA190747 391 bp mRNA EST 15-JAN-1997
DEFINITION zpr8d05.r1 Stratogene HeLa cell s3 937216 Homo sapiens CDNA clone IMAGE:627273 5' similar to SW-PIB.HUMAN P42338 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;,
mRNA sequence.
ACCESSION AA190747
MIMD 91779133
VERSION AA190747.1 GI:1779133
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 391)
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chipelli,B., Chissoe,S., Dietrich,N., Dubnue,T., Favellio,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,

TITLE
JOURNAL
MEDLINE

COMMENT

On May 18, 1995 this sequence version replaced gi:810974.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LINT ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -28413 rev2 from Amersham
High quality sequence stop: 233.

FEATURES
Source

1. 391
/organism="Homo sapiens"
/db_xref="GDB:5045971"
/db_xref="taxon:9606"
/clone="IMAGE:627273"
/clone_1lp="Scratagene Hela cell s3 937216"
/sex="female"
/dev_stage="Hela S3 cell line"
/lab_host="SOLR (Kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. Hela S3
epithelioid carcinoma cells grown to semi-confluency
without induction. Average insert size: 1.5 kb; Uni-ZAP
Vector. -5' adaptor sequence: 5' GAATTCGACGACAG 3' -3'
adaptor sequence: 5' CTCGACGATTTTATTTTATTT 3' "

BASE COUNT
109 a 79 c 89 g 109 t 5 others

ORIGIN

Query Match 70.9%; Score 24.8; DB 29; Length 391;
Best Local Similarity 76.7%; Pred. No. 1.7;
Matches 23; Conservative 5; Mismatches 2; Indels 0; Gaps 0.

OY 6 TGGCCACTGTTTCATATGATTTTGCCA 35
|||||:||||:||||:|||||
Db 183 TGGCCAGCTCTTCACATGACTTGGACA 212

RESULT 5
N76563
LOCUS

DEFINITION
N76563 479 bp mRNA EST 28-JAN-1997
IMAGE:245164 5' similar to SW-PLIB HUMAN P42338
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;
mRNA sequence.
N76563
g1239141
N76563.1 GI:1239141
EST.

ACCESSION
NID
VERSION
KEYWORDS
SOURCE

ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiparelli,B.,
Chisose,S., Dietrich,N., Dubuque,T., Favellio,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Maddis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rolling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.

TITLE
JOURNAL
Genome Res. 6 (9), 807-828 (1996)

Generation and analysis of 280,000 human expressed sequence tags

MEDLINE 97044478
 COMMENT On May 5, 1995 this sequence version replaced gi:798462.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 *Insert length: 1532 Std Error: 0.00
 Seq primer: reverse ET
 High quality sequence stop: 368.
 Location/Qualifiers

FEATURES

Source

1. .479
 /organism="Homo sapiens"
 /db_xref="GDB:3794410"
 /db_xref="taxon:9606"
 /map="21"
 /clone="IMAGE:245164"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 1st strand cDNA was primed with a Pac I - oligo(dT) primer
 [5' AGCTGAGAGATTAATTAAGATCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 120 a 99 c 111 g 144 t 5 others
 ORIGIN

Query Match 64.0%; Score 22.4; DB 25; Length 479;
 Best Local Similarity 76.9%; Pred. No. 17;
 Matches 20; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 7 GGCCARCTGTTCAATGATGATTTGG 32
 |||||:||||:||||:||||:||||:||||
 Db 359 GGCCAGCTCTTCACATTCATCTTGG 384

RESULT 6
 A1155136 447 bp mRNA EST 30-SEP-1998
 LOCUS U699f06.r1 Soares mouse uterus NMPu Mus musculus cDNA clone
 DEFINITION IMAGE:1450211 5' similar to SW:P11A.MOUSE P42337
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
 mRNA sequence.

ACCESSION A1155136
 NID 93683605
 VERSION A1155136.1 GI:3683605

SOURCE

ORGANISM Homo sapiens
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 447)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Apr 18, 1995 this sequence version replaced gi:775514.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:923527
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28n13 rev2 ET from Amersham
 High quality sequence stop: 438.
 Location/Qualifiers

FEATURES

Source

1. .447
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="IMAGE:1450211"
 /clone_idb="Soares mouse uterus NMPu"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; 1st strand cDNA was prepared from
 pregnant mouse uterus, and was then primed with a Not I -
 oligo(dT) primer. Double-stranded cDNA was ligated to Eco
 RI adaptors (Pharmacia), digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library is normalized. Library was constructed by
 Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 139 a 79 c 101 g 127 t 1 others
 ORIGIN

Query Match 61.7%; Score 21.6; DB 42; Length 447;
 Best Local Similarity 70.0%; Pred. No. 36;
 Matches 21; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 6 TGCCARCTGTTCAATGATGATTTGGCCA 35
 |||||:||||:||||:||||:||||:||||
 Db 21 TTGAAGCTGTTTCATTAATGATTTGGCCA 50

RESULT 7
 A1193439 378 bp mRNA EST 19-MAY-1997
 LOCUS z141D09.s1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:665945
 DEFINITION 3', mRNA sequence.

ACCESSION A1193439
 NID 91782867
 VERSION A1193439.1 GI:1782867

KEYWORDS

EST.

SOURCE

ORGANISM Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 378)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,
 Chissole, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Marra, M., Moore, B., Morris, M., Parsons, J., Prange, C., Rickin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
 and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810


```

source
1.509
/organism="Homo sapiens"
/db_xref="GDB:408627"
/db_xref="taxon:9606"
/map="9"
/clone="IMAGE:36126"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site:1: Not
I; Site:2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5];
ACTGGAAGAAATTCGCGCCGACAGAAATTTTCTTTTCTTTT 3');
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      134 a      104 c      116 g      143 t      12 others
ORIGIN

Query Match
Best Local Similarity 76.9%; Score 21.2; DB 21; Length 509;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GGCACATGTTTCATATGACTTTGG 32
||||| |||||:||||:|||||
db 87 GGCACATGTTTCATATGACTTTGG 112

RESULT 10
LOCUS      R22699      330 bp      mRNA      EST      19-APR-1995
DEFINITION YG08e01.1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:31318 5' similar to SP:P100 BOVIN P32871 PHOSPHATIDYLINOSITOL
3-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.
R22699      92662210
NID          92662210
VERSION      R22699.1 GI:777456
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 330)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaskis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

*Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 1762
High quality sequence stops: 271 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1762 Std Error: 0.00
Seq primer: M13rp1
High quality sequence stop: 271.
Location/Qualifiers
1..330
/organism="Homo sapiens"
/db_xref="GDB:40365"
/db_xref="taxon:9606"

```

```

/clone="IMAGE:31318"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site:1: Not
I; Site:2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5];
ACTGGAAGAAATTCGCGCCGACAGAAATTTTCTTTTCTTTT 3');
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT      92 a      62 c      77 g      99 t
ORIGIN

Query Match
Best Local Similarity 76.9%; Score 21.2; DB 22; Length 330;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 7 GGCACATGTTTCATATGACTTTGG 32
||||| |||||:||||:|||||
db 87 GGCACATGTTTCATATGACTTTGG 112

RESULT 11
LOCUS      AB009107      574 bp      mRNA      EST      05-DEC-1997
DEFINITION AB009107 Chlamydomonas W80 lambda ZAP II Chlamydomonas sp. cDNA
similar to phosphatidylinositol 3-kinase, mRNA sequence.
AB009107
ACCESSION   AB009107
NID          92662210
VERSION      AB009107.1 GI:2662210
KEYWORDS     EST.
SOURCE       Chlamydomonas sp.
ORGANISM     Chlamydomonas sp.
REFERENCE    Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
AUTHORS      Chlamydomonadaceae; Chlamydomonas.
1 (bases 1 to 574)
Kanaboshi,H., Ikeda,K. and Miyasaka,H.
Isolation of several anti-stress genes from the halotolerant green
alga Chlamydomonas by a simple functional expression screening in
E. coli
Unpublished (1998)
On Sep 19, 1997 this sequence version replaced gi:1520267.

JOURNAL
COMMENT

*Contact: Hitoshi Miyasaka
Technical Research Center
The Kansai Electric Power Co.
Nakoji 3-Chome 11-20, Amagasaki, Hyogo 661, Japan
Tel: 81-6-494-9840
Fax: 81-6-494-9728
Email: hmiya@mb.infoweb.or.jp.

TITLE
JOURNAL
COMMENT

FEATURES
source
1..574
Location/Qualifiers
/organism="Chlamydomonas sp."
/strain="W80"
/db_xref="taxon:3056"
/clone_lib="Chlamydomonas W80 lambda ZAP II"
/note="Vector: lambda ZAP II; The cDNA clone was isolated
from the halotolerant green alga Chlamydomonas W80 by a
functional expression screening in E. coli cells. The
principle of the screening method was based on the
acquisition of stress tolerance of the bacterial cells
carrying the cDNA."

BASE COUNT      96 a      197 c      184 g      97 t
ORIGIN

Query Match
Best Local Similarity 71.9%; Score 21.2; DB 37; Length 574;
Matches 20; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```


GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:15 ; Search time 539.84 Seconds

(without alignments)
135,498 Million cell updates/sec

Title: US-09-325-095-15

Perfect score: 23

Sequence: 1 AATTGACACACTGGCATGCCGAT 23

Scoring table: IDENTITY_NDC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_to.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*
16: gb_un.*
17: gb_vl.*
18: em_kun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_ov.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_to.*
30: em_sts.*
31: em_sy.*
32: em_un.*
33: em_vl.*
34: gb_htg1.*
35: gb_htg2.*
36: gb_in1.*
37: gb_in2.*
38: em_ba1.*
39: em_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	23	100.0	3207	3	BOVPHOS3KN	M93252 Bovine phop

C	2	23	100.0	3498	5	A37232	A37232 Sequence 1
C	3	23	100.0	225	5	A37243	A37243 Sequence 12
C	4	23	100.0	3207	10	HSU79143	U79143 Human phosp
C	5	21.4	93.0	3424	10	HSPH13K	229090 H.sapiens m
C	6	21.4	93.0	3207	12	GMU03279	U03279 Mus muscitu
C	7	18.2	79.1	640	14	G30932	G30932 pig STS CP-
C	8	17.2	74.8	6502	11	AC003691	AC003691 Human Chr
C	9	17.2	74.8	5085	12	MMANREC	211974 M.musculus
C	10	17.2	74.8	190000	35	AC006427	AC006427 Homo sapi
C	11	16.8	73.0	6067	42	AC007766	AC007766 Homo sapi
C	12	16.8	73.0	18483	2	AF005383	M34459 C.saccharol
C	13	16.8	73.0	2062	5	184210	AF005383 Caldicell
C	14	16.8	73.0	164	5	184217	184210 Sequence 8
C	15	16.8	73.0	95214	8	ATAC06526	184217 Sequence 19
C	16	16.8	73.0	110608	9	HS283E3	AC006526 Arabidops
C	17	16.8	73.0	107418	9	HS51E13	AL031282 Human DNA
C	18	16.8	73.0	5586	12	D67015	AL022163 Human DNA
C	19	16.8	72.2	9117	1	PEMETH	D67015 Mouse mRNA
C	20	16.8	72.2	10763	5	AR008973	L39364 Porcine res
C	21	16.8	73.0	3292	17	PPSSG1A	L39369 Porcine res
C	22	16.8	73.0	2050	17	PRU03040	U03040 Porcine rep
C	23	16.8	73.0	167439	35	AC007299	AC007299 Drosophila
C	24	16.6	72.2	9117	1	PEMETH	L29642 Pseudomonas
C	25	16.6	72.2	2179	3	CAPURA	AF079548 Ovis arie
C	26	16.6	72.2	10763	5	AR008973	AR008973 Sequence
C	27	16.6	72.2	5559	5	166429	166429 Sequence 3
C	28	16.6	72.2	5559	5	173354	173354 Sequence 3
C	29	16.6	72.2	5559	5	173361	173361 Sequence 3
C	30	16.6	72.2	5559	5	181321	181321 Sequence 3
C	31	16.6	72.2	104364	8	ATAC006919	AC006919 Arabidops
C	32	16.6	72.2	114929	9	AP000050	AP000050 Homo sapi
C	33	16.6	72.2	151516	9	AP000117	AP000117 Homo sapi
C	34	16.6	72.2	100000	9	AP000193	AP000193 Homo sapi
C	35	16.6	72.2	78539	9	AP000313	AP000313 Homo sapi
C	36	16.6	72.2	167514	9	HS242N11	AL023655 Human DNA
C	37	16.6	72.2	154902	11	AC005799	AC005799 Homo sapi
C	38	16.6	72.2	369	11	HSTSN3	AF064247 Homo sapi
C	39	16.6	72.2	5366	17	HAPVX	X02449 Hamster pap
C	40	16.6	72.2	5366	17	PPCCGAAA	M26281 Hamster pap
C	41	16.6	72.2	133572	35	AC006965	AC006965 Homo sapi
C	42	16.6	72.2	39478	36	CEF53E1	281088 Caenorhabdi
C	43	16.6	72.2	34190	36	CEK04G11	278544 Caenorhabdi
C	44	16.6	72.2	91019	37	AC001658	AC001658 Drosophila
C	45	16.6	72.2	175775	42	AC006947	AC006947 Homo sapi

ALIGNMENTS

RESULT 1
BOVPHOS3KN/C
LOCUS
DEFINITION Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds
ACCESSION M93252
NTD 9163519
VERSION M93252.1 GI:163519
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE Bos taurus cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R., Panayotou, G., Ruiz-Iarrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D. Phosphatidylinositol 3-kinase: Structure and expression of the 110 kd catalytic subunit
JOURNAL MEDLINE
CELL 70, 419-429 (1992)
92354059
FEATURES
SOURCE location/Qualifiers
1..3207

CDS

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AA30698.1"
/db_xref="PDB:1g63520"
/db_xref="GI:163520"
/translation="MPRRSSGELMGHIMPRRIIVCELLPNGMIVTTECLREATLIT
IKHELFKARKYPLHOLLQDESSYIFVSTQAEAREEFEDETRRLDCLPFPFKVI
EYVGNREKILNREIGFATGMPVCEFDMDKPEVODERRNIIINCKEAVDLNLSPI
SRAMVYPPNVSSPELPHKTIYKDKGQIIYIVIVTSPNNDKQYITLKHNDVPE
OYIAEIRKTRSMLSSEQLKCVLEYOGKYLKCGDEFFLEKYPISQYKIRSC
IMGRPNLMIMAKESLSQLPMDCTFMSYSRISTATPYNGETSTSLVINSAL
RIKICATVYNNIRIDIKIYVGTGIYHGGELCDVNTQVPCSPNPNENLNDIY
IPDLPRARLCLISGKRGKAKKEHCPANGNINLFPYDTLVSGKALNMPVPH
GLEDLNPIGTGSNPNKEPCLLEEFWFSYVKFPDMSVIEEHANNSVREAGTSY
SHAGLSNRLARNELRENDKQRLAICTRDPLSETTEDEKDFLMSHRHYCVITPEILP
KLLSVKNRSRDEVAQMTCLVMDPPIKPEQAMELDQNPDPVNRGFAVCLERYLT
DRLSQYLIQVQVLYKQYLDNLVRFLLKALNQRIGHEFFWHLKSEMNKTVSQ
REGLLSEYCRAGMYLKLNRQVAMEKLINLTLKQEKDEQKQYMKFLVEQMR
RPEDMDALQGLSPINPAHQNLRLBECRIMSASRRPLMNENPDIMSELRHONNE
IIFKNGDDLRODMLTQIIRIMENIMQNGDLRLMPCGCLSGDCVGLIEVNRSH
IMQIOCKGGLGALQFNSHTLHQMLKDNKKEITDVAIDLETFRSCAGVAFILIGIG
DHRNNSINVKDQGFHIDFGHFLDKRKRKRGKRVFVLTQDFLIVISGADECT
KTRFEREOECYKAYLAIRQHANLEFINLSMGLSGMPELQSFDDIAYIRRTALDK
TEQEALEYFMKQMDAHGGMWTKMDWIFHTIKOHALN"

BASE COUNT 1028 a 581 c 680 g 918 t

ORIGIN

Query Match 100.0%; Score 23; DB 3; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
|||||
Db 383 AATTCACACACTGGCATGCCGAT 361

RESULT 2
LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
NID 92294345
VERSION A37232.1 GI:2294345
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 1 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source location/Qualifiers
1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"

CDS
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PDB:E304808"
/db_xref="PDB:92294346"
/db_xref="GI:2294346"
/translation="MPRRSSGELMGHIMPRRIIVCELLPNGMIVTTECLREATLIT
IKHELFKARKYPLHOLLQDESSYIFVSTQAEAREEFEDETRRLDCLRPQPLAKVI

EYVGNREKILNREIGFATGMPVCEFDMDKPEVODERRNIIINCKEAVDLNLSPI
SRAMVYPPNVSSPELPHKTIYKDKGQIIYIVIVTSPNNDKQYITLKHNDVPE
OYIAEIRKTRSMLSSEQLKCVLEYOGKYLKCGDEFFLEKYPISQYKIRSC
IMGRPNLMIMAKESLSQLPMDCTFMSYSRISTATPYNGETSTSLVINSAL
RIKICATVYNNIRIDIKIYVGTGIYHGGELCDVNTQVPCSPNPNENLNDIY
IPDLPRARLCLISGKRGKAKKEHCPANGNINLFPYDTLVSGKALNMPVPH
GLEDLNPIGTGSNPNKEPCLLEEFWFSYVKFPDMSVIEEHANNSVREAGTSY
SHAGLSNRLARNELRENDKQRLAICTRDPLSETTEDEKDFLMSHRHYCVITPEILP
KLLSVKNRSRDEVAQMTCLVMDPPIKPEQAMELDQNPDPVNRGFAVCLERYLT
DRLSQYLIQVQVLYKQYLDNLVRFLLKALNQRIGHEFFWHLKSEMNKTVSQ
REGLLSEYCRAGMYLKLNRQVAMEKLINLTLKQEKDEQKQYMKFLVEQMR
RPEDMDALQGLSPINPAHQNLRLBECRIMSASRRPLMNENPDIMSELRHONNE
IIFKNGDDLRODMLTQIIRIMENIMQNGDLRLMPCGCLSGDCVGLIEVNRSH
IMQIOCKGGLGALQFNSHTLHQMLKDNKKEITDVAIDLETFRSCAGVAFILIGIG
DHRNNSINVKDQGFHIDFGHFLDKRKRKRGKRVFVLTQDFLIVISGADECT
KTRFEREOECYKAYLAIRQHANLEFINLSMGLSGMPELQSFDDIAYIRRTALDK
TEQEALEYFMKQMDAHGGMWTKMDWIFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t

ORIGIN

Query Match 100.0%; Score 23; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 0.08;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
|||||
Db 383 AATTCACACACTGGCATGCCGAT 361

RESULT 3
LOCUS A37243 23 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 12 from Patent WO9403609.
ACCESSION A37243
NID 92294354
VERSION A37243.1 GI:2294354
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 23)
AUTHORS Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 12 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source location/Qualifiers
1..23
/organism="unidentified"
/db_xref="taxon:32644"
7 a 7 c 4 g 5 t

BASE COUNT 7 a 7 c 4 g 5 t

ORIGIN

Query Match 100.0%; Score 23; DB 5; Length 23;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCCGAT 23
|||||
Db 1 AATTCACACACTGGCATGCCGAT 23

RESULT 4
LOCUS HSU79143 3207 bp mRNA PRI 02-JAN-1997
DEFINITION Human phosphoinositide 3'-hydroxylase p110-alpha subunit mRNA,
complete cds.
ACCESSION U79143
NID 91763625
VERSION U79143.1 GI:1763625

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS 1 (bases 1 to 3207)
TITLE Olfiff, A. and Heilmbrock, D.C.
Catalytic Activity of the p110-alpha Subunit of Human
Phosphoinositide 3'-Hydroxylase is Required for Signal
Transduction
JOURNAL Bloorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Stirdivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M.,
Olfiff, A. and Heilmbrock, D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
Sumneytown Pike, West Point, PA 19486, USA
FEATURES
source
location/Qualifiers
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/codon_start=1
/product="phosphoinositide 3'-hydroxylase p110-alpha
subunit"
/protein_id="AAB39753.1"
/db_xref="PID:91763626"
/db_xref="GI:1763626"
/translation="MPPRSSGELMGTHLMPRIIVLVECLLPNGMIVTECREATLV
IKHLEFKARKYPIHQLDDESYIFSVTOEAREEFEDTRICDLRLPPLKVI
EPVGNREKILNREIGFALGMPVEFPMWVNDPEODERNILNVCKEAVDRLDSPH
SRATVYVPPHVSSELPKHYNLKDLGQIIVYVIVSPNNDOKYTLKINHCVE
QVIAEAIKRTKRSMLSEQLKLCVLEQGYILKVCDCDEFLKPLSOYKTRSC
IMGRMNLMMAKESLYSOLPMDCTFMPYSRISATPYMNGETSKRLMVLNRL
RIKILCATYVNLVNDIKIVRTGIGHGEPICLDNVTQVPCSNPMMNMLNDIY
IPDLPRARCLSCISYKGRKGAECHECPPLMNGINLEFDITDILVSGMALNLPVPH
GLEDLNPIGVTSNPKETPCLEDFDMSVVKFPMDSVIEHANSVSREGFSY
SHAOLSKRLANDNELRENDKQLAISTRDPLSETTEKDFLMSHRHYCTIPEILP
KLLSVKNSDEVAQWYCLVNDMPPIKEQAMELLDCNYPDPWVGRGAVCLERKYL
DRLSOYLIOVLYKTEQYLDNLVLEFLKALTNORIGHEFWHLKSEHNNTVSQ
RPFMDALOGFLSPNPAHQGNRLRECRIMSSAKRPLMLNMPNPDIMSLLFQNSH
IIFKNGDLRODMTLTQIRIMENIMONGIDLRLPYGCLSIDCVGLLIVVNSHT
IMOJCKGKLGALQFNSHTLHQLKDKNKEIYDAIDFTBSCAGCVATFTLGIG
DRHNSNINVKDQGLFHDHFKKKKRGYRERPEVLTQDFLIVISGAOECT
KTRFEEROEMCYAAYLAIRQHANLFTNLESMGSGMPELQSFDDIAYIKRTALDK
TEOALTEYFMQNDAHGCGTTKMDTFHTIKQHALN"

BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN

Query Match 100.0%; Score 23; DB 10; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.08; Mismatches 0; Indels 0; Gaps 0;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AATTCACACTGCGATGCCGAT 23
|||||
DB 383 AATTCACACTGCGATGCCGAT 361

RESULT 5
HSPH13K 3424 bp mRNA PRI 24-AUG-1995
LOCUS H.sapiens mRNA for phosphatidylinositol 3-kinase.
DEFINITION 229090
ACCESSION 9472990
VERSION 229090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

REFERENCE Primates; Catarrhini; Homnidae; Homo.
AUTHORS 1 (bases 1 to 1068)
TITLE Volinia, S., Hiles, I., Ormondcroft, E., Mizetic, D., Antonacci, R.,
Rocchi, M. and Waterfield, M.D.
Molecular cloning, cDNA sequence, and chromosomal localization of
the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL Genomics 24 (3), 472-477 (1994)
REFERENCE 95229146
AUTHORS 2 (bases 1 to 3424)
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK
FEATURES
source
location/Qualifiers
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CAA82333.1"
/db_xref="PID:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPPRSSGELMGTHLMPRIIVLVECLLPNGMIVTECREATLV
IKHLEFKARKYPIHQLDDESYIFSVTOEAREEFEDTRICDLRLPPLKVI
EPVGNREKILNREIGFALGMPVEFPMWVNDPEODERNILNVCKEAVDRLDSPH
SRATVYVPPHVSSELPKHYNLKDLGQIIVYVIVSPNNDOKYTLKINHCVE
QVIAEAIKRTKRSMLSEQLKLCVLEQGYILKVCDCDEFLKPLSOYKTRSC
IMGRMNLMMAKESLYSOLPMDCTFMPYSRISATPYMNGETSKRLMVLNRL
RIKILCATYVNLVNDIKIVRTGIGHGEPICLDNVTQVPCSNPMMNMLNDIY
IPDLPRARCLSCISYKGRKGAECHECPPLMNGINLEFDITDILVSGMALNLPVPH
GLEDLNPIGVTSNPKETPCLEDFDMSVVKFPMDSVIEHANSVSREGFSY
SHAOLSKRLANDNELRENDKQLAISTRDPLSETTEKDFLMSHRHYCTIPEILP
KLLSVKNSDEVAQWYCLVNDMPPIKEQAMELLDCNYPDPWVGRGAVCLERKYL
DRLSOYLIOVLYKTEQYLDNLVLEFLKALTNORIGHEFWHLKSEHNNTVSQ
RPFMDALOGFLSPNPAHQGNRLRECRIMSSAKRPLMLNMPNPDIMSLLFQNSH
IIFKNGDLRODMTLTQIRIMENIMONGIDLRLPYGCLSIDCVGLLIVVNSHT
IMOJCKGKLGALQFNSHTLHQLKDKNKEIYDAIDFTBSCAGCVATFTLGIG
DRHNSNINVKDQGLFHDHFKKKKRGYRERPEVLTQDFLIVISGAOECT
KTRFEEROEMCYAAYLAIRQHANLFTNLESMGSGMPELQSFDDIAYIKRTALDK
TEOALTEYFMQNDAHGCGTTKMDTFHTIKQHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 93.0%; Score 21.4; DB 10; Length 3424;
Best Local Similarity 95.7%; Pred. No. 0.53; Mismatches 1; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AATTCACACTGCGATGCCGAT 23
|||||
DB 395 AATTCACACTGCGATGCCGAT 373

RESULT 6
MMU03279/c 3207 bp mRNA ROD 25-MAY-1994
LOCUS Mus musculus Balb/c phosphatidylinositol 3-kinase 110 kDa subunit
DEFINITION mRNA, complete cds.
ACCESSION U03279
VERSION 9414994
KEYWORDS U03279.1 GI:414994
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
1 (bases 1 to 3207)

AUTHORS Klippel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
TITLE The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE 94187738
REFERENCE 2 (bases 1 to 3207)
AUTHORS Klippel-Giese, A.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dalich Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
FEATURES
 source
 1..3207
 /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"
 1..3207
 /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="PID:q41495"
 /db_xref="GI:41495"
 /translation="MPPRSSEGLMGIIHMPRIIVECLLPNGMIVTECLREATV
 IKHELFRARKYPIHQLLODEPTSYFVSVOAEAREEFEDTRRLCDLRLPPLKVI
 EYGNREKILNREIFVIGMPCEDMVKDPEVDFFRNILNCKEANDLIDNSPH
 SRAMVYPPNVSSEPELPKHIITKLDKGIIVIVIVSPPNDKQITLKNDCVPE
 QVIAEIRKTRSMLESEQLKCYLEYOGKYLKCGDEXFLEKYSQYKIRSC
 IMGRPNLMMAKESLYSQLIDSETPMSRSRISTATPYNGESTKSLVINSAL
 RIKLCATVYNNIRIDIKIVRTGIGHGEPCDVTORVPCSNPRNEMVNDIY
 IDPLRLARLCLSCISGKRAKEEHLANGNINLFPYTDIVSGKALNMPVPH
 GLEDLNPICGASNPKEPCLTEPEPSSVYKPRDMSVIEAHMSVSEAGPSY
 SHTGJSNRLARNELENDKEQRLCTDPISEITGEKDLMSRHRCVITPELP
 KLLSVKNMSRDEVAQMYCLVDWPIKEQAMELIDCNYPDPVASEPVRLELYTQ
 DKLSQYLIVQVLYKEQYLDNLVRFLLKRALTNQRIGHEFWHLKSEMKRYVQ
 REGLLSEYCRACGMYIKLNQVEMEKILNLTDLKQEKDEQKVMKFLVDMR
 ODFMDALGFLSPNPAHQLNLRLECRIMSARPLMLNEMNDIMSELFPONE
 ILEKNGDDLRDMQTLQITRIMENIMONGDILPMIPYCLSLGDCVGLIYVRSHT
 IMQIOCKGLKALQFNHSHHTLQWLNKRGELTYDAIDLETRSCAGTCVATITIG
 DRNSNINVKDQGLFHIDGFLHKKRKFYKRRAPVFLVLDPLIVISGAEVET
 KTRFEREEMCYKAYLAIRAHANLEINFSMLSGMPELQSPFDIAYIRKTLALDK
 TEQEALVETFKOMNDAHGGMTKMDIMFHTIKQHLN"

CDS
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="PID:q41495"
 /db_xref="GI:41495"
 /translation="MPPRSSEGLMGIIHMPRIIVECLLPNGMIVTECLREATV
 IKHELFRARKYPIHQLLODEPTSYFVSVOAEAREEFEDTRRLCDLRLPPLKVI
 EYGNREKILNREIFVIGMPCEDMVKDPEVDFFRNILNCKEANDLIDNSPH
 SRAMVYPPNVSSEPELPKHIITKLDKGIIVIVIVSPPNDKQITLKNDCVPE
 QVIAEIRKTRSMLESEQLKCYLEYOGKYLKCGDEXFLEKYSQYKIRSC
 IMGRPNLMMAKESLYSQLIDSETPMSRSRISTATPYNGESTKSLVINSAL
 RIKLCATVYNNIRIDIKIVRTGIGHGEPCDVTORVPCSNPRNEMVNDIY
 IDPLRLARLCLSCISGKRAKEEHLANGNINLFPYTDIVSGKALNMPVPH
 GLEDLNPICGASNPKEPCLTEPEPSSVYKPRDMSVIEAHMSVSEAGPSY
 SHTGJSNRLARNELENDKEQRLCTDPISEITGEKDLMSRHRCVITPELP
 KLLSVKNMSRDEVAQMYCLVDWPIKEQAMELIDCNYPDPVASEPVRLELYTQ
 DKLSQYLIVQVLYKEQYLDNLVRFLLKRALTNQRIGHEFWHLKSEMKRYVQ
 REGLLSEYCRACGMYIKLNQVEMEKILNLTDLKQEKDEQKVMKFLVDMR
 ODFMDALGFLSPNPAHQLNLRLECRIMSARPLMLNEMNDIMSELFPONE
 ILEKNGDDLRDMQTLQITRIMENIMONGDILPMIPYCLSLGDCVGLIYVRSHT
 IMQIOCKGLKALQFNHSHHTLQWLNKRGELTYDAIDLETRSCAGTCVATITIG
 DRNSNINVKDQGLFHIDGFLHKKRKFYKRRAPVFLVLDPLIVISGAEVET
 KTRFEREEMCYKAYLAIRAHANLEINFSMLSGMPELQSPFDIAYIRKTLALDK
 TEQEALVETFKOMNDAHGGMTKMDIMFHTIKQHLN"

BASE COUNT 958 a 693 c 734 g 822 t

ORIGIN

Query Match 93.0%; Score 21.4; DB 12; Length 3207;
Best Local Similarity 95.7%; Pred. No. 0.53;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AATTACACACTGGCATGCCGAT 23
 ||||||||||||||||||||
DB 383 AATTACACACTGGCATGCCAAT 361

RESULT 7
LOCUS G30932 640 bp DNA STS 15-NOV-1996
DEFINITION pig STS CP-000R, sequence tagged site.
ACCESSION G30932
NID 91669818
VERSION G30932.1 GI:1669818
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE pig strain-Melishan Pig genomic DNA was prepared by standard procedure.
ORGANISM Sus scrofa
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Cetartiodactyla; Suidae; Sus.
AUTHORS 1 (bases 1 to 640)
TITLE Comparative gene mapping between human and pigs
JOURNAL Unpublished (1996)
COMMENT

Contact: Sun, H. S.
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150
 Tel: 515-294-4209
 Fax: 515-294-2401
 Email: hssun@iastate.edu

Primer A: CTAGTCCCTGCTATTGGC
Primer B: TCTTTGGGAGACGTCATTC
STS size: 640
PCR Profile:

Presack: 95 degree C for 2 minutes
Denaturation: 95 degree C for 0.5 minute
Annealing: 58 degree C for 1 minute
Polymerization: 72 degree C for 1 minutes
PCR cycles: 32
Thermal Cycler: MJ Research

Protocol:
 Template: 30-100 ng
 Primers: 0.3 uM
 dNTPs: each 200 uM
 Taq Polymerase: 0.033 units/uL
 Total Vol: 15 uL

Buffer:

MgCl2: 1.5 mM
 KCl: 50 mM
 Tris-HCl: 10 mM
 pH: 8.3.

FEATURES
 source
 1..640
 Location/Qualifiers
 /organism="Sus scrofa"
 /db_xref="taxon:9823"
 /strain="Melishan"

STS
 >1..<640
 /gene="CP"
 <1..>640
 /gene="CP"
 /note="ceruloplasmin"

BASE COUNT 186 a 126 c 142 g 186 t

ORIGIN

Query Match 79.1%; Score 18.2; DB 14; Length 640;
Best Local Similarity 87.0%; Pred. No. 27;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 AATTACACACTGGCATGCCGAT 23
 ||||||||||||||||||||
DB 292 AATTACACACTGGCATCTAAT 314

RESULT 8
LOCUS AC003691 65020 bp DNA PRI 29-OCT-1998
DEFINITION Human Chromosome 11q23 PAC clone pD254e13, complete sequence.
ACCESSION AC003691
NID 93808083
VERSION AC003691.1 GI:3808083
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Evans, G.A., Athanasiou, M., Aguayo, P., Armstrong, D., Basil, M.,
 Buettner, T., Bumsister, R., Card, P., Desaliboat, F., Dunn, J.,
 English, C., EhrIDGE, S., Garner, H.R., Gee, V., Gordon, M., Gotway, G.,
 Grant, O., Hamner, L., Joslin, J., Lewis, E., Loo, H., Loo, K.N.,
 Major, T., McFarland, J., Newton, J., Osborne-Lawrence, S.,
 Schageman, J., Schultz, R.A., Stimson, S., Syed, M. and Ward, T.
TITLE HTGS Submission

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 65020)
AUTHORS Evans,G.A., Athanasiou,M., Basil,M., Bradbury,P., Brinac,S., Bunesler,R., Davis,C., English,C., Franklin,T.L., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Haber,L., Harris,J., Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.

TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-8591, USA
REFERENCE 3 (bases 1 to 65020)
AUTHORS Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basil,M., Buettner,J., Butler,C., Card,P., desaiiboa,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Haber,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Walker,K. and Ward,T.

TITLE Direct Submission
JOURNAL Submitted (29-OCT-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235-8591, USA

COMMENT On Oct 29, 1998 this sequence contains the entire insert of clone PDJ254e13. PDJ254e13 comes from the RPOC-3 PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.85. In addition, attempts have been made to assure 98% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.

FEATURES
source 1..65020
location/Qualifiers
db_xref="taxon:9606"
/organism="Homo sapiens"
repeat_region complement(5..86)
/rpt_family="THE1"
repeat_region 1040..1347
/rpt_family="Alu"
repeat_region 1543..1632
/rpt_family="THE1"
repeat_region complement(2645..2925)
/rpt_family="Alu"
repeat_region complement(3404..3692)
/rpt_family="Alu"
repeat_region 5568..5990
/rpt_family="Alu"
repeat_region complement(6925..7021)
/rpt_family="MIR"
repeat_region 9109..9172
/rpt_family="MLT1"
repeat_region complement(9452..9630)
/rpt_family="Alu"
repeat_region complement(10004..10079)
/rpt_family="MLT1"
repeat_region complement(10087..10280)
/rpt_family="MLT1"
repeat_region complement(10686..10959)
/rpt_family="Alu"
repeat_region complement(12119..12387)
/rpt_family="Alu"
repeat_region complement(15054..15256)
/rpt_family="Alu"
repeat_region complement(15762..16055)
/rpt_family="MIR4.2"
repeat_region complement(16280..16476)
/rpt_family="L1"
repeat_region 17758..17984
/rpt_family="Alu"
repeat_region complement(18051..18258)

repeat_region /rpt_family="L1"
20160..20437
repeat_region /rpt_family="Alu"
22430..22738
repeat_region /rpt_family="Alu"
26073..26379
repeat_region /rpt_family="Alu"
complement(27905..28181)
repeat_region /rpt_family="Alu"
30140..30466
repeat_region /rpt_family="Alu"
32260..32543
repeat_region /rpt_family="Alu"
34389..34674
repeat_region /rpt_family="Alu"
complement(38169..38537)
repeat_region /rpt_family="THE1"
39230..39318
repeat_region /rpt_family="MIR"
40108..40247
repeat_region /rpt_family="MIR"
complement(40311..40580)
repeat_region /rpt_family="Alu"
40618..40714
repeat_region /rpt_family="L1"
46149..46227
repeat_region /rpt_family="MLT1"
47776..48090
repeat_region /rpt_family="Alu"
complement(51040..51183)
repeat_region /rpt_family="MIR5"
complement(52705..52989)
repeat_region /rpt_family="Alu"
complement(53823..53944)
repeat_region /rpt_family="MIR"
54921..55201
repeat_region /rpt_family="Alu"
complement(55486..55783)
repeat_region /rpt_family="Alu"
60173..60468
repeat_region /rpt_family="Alu"
BASE COUNT 20406 a 11757 c 12727 g 20130 t
ORIGIN

Query Match 74.8%; Score 17.2; DB 11; Length 65020;
Best Local Similarity 86.4%; Pred. No. 58;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 ATTACACACTGGCATGCCGAT 23
DB 61843 ATTACACACAGCATGCACAT 61864

RESULT 9
MMANNREC 5085 bp mRNA ROD 25-SEP-1996
LOCUS M.musculus mRNA for macrophage mannose receptor.
DEFINITION 211974
ACCESSION 552997
NTID 211974.1 GI:52997
VERSION macrophage mannose receptor; mannose receptor.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 4467)
AUTHORS Harris,N., Super,M., Rits,M., Chang,G. and Ezekowitz,R.A.
TITLE Characterization of the murine macrophage mannose receptor: demonstration that the downregulation of receptor expression mediated by interferon-gamma occurs at the level of transcription

JOURNAL Blood 80 (9), 2363-2373 (1992)
MEDLINE 93043353

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS Lamerdin,J.E., McCreedy,P.M., Richardson,P., Burkhardt-Schultz,K., Gordon,L., Stillwagen,S., Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Sakaladis,G., Dangnan,L., Ertler,A., Christensen,M., Georgescu,A., Avila,J., Liu,S., Altix,C., Andeise,T., Scott,D., Amico-Keller,G., Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G., Kronmiller,B., Arellano,A., Sanders,C., Ow,D., Nolan,M., Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
TITLE Sequence analysis of a 6.4 kb region in 19p13.3 between CDC34 and VAV1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 39838)
AUTHORS Lamerdin,J.E.
JOURNAL Direct Submission
COMMENT Submitted (09-JUN-1999) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
Map and sequence oriented from P telomere to centromere. Cosmid R26510 overlaps cosmid F25540 (AC006277) to the left from bases 1 to 2,676 of this accession, and overlaps cosmid R31236 from bases 30,971 to 39,838. Additional map and sequence information may be obtained at: <http://www-bio.llnl.gov/dbp/genome/genome.html>.

FEATURES
source
1..39838
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R26510"
/chromosome="19"
/map="19p13.3 between CDC34 and VAV1"
/cell_line="SHL2-B"
/clone_id="TL19NC03 R chromosome 19-specific cosmid library"
/note="Cosmid library constructed at LLNL from flow-sorted chromosomes from human-hamster hybrid 5H2-B, which carries chromosome 19 as its only human chromosome."
1..159
/note="BLASTN similarity to 257824 (98..256): match: 0.98, score: 5.9e-55; database searched: nt; H.sapiens CPG island DNA genomic MseI fragment, clone 198b12, reverse read cp9198b12.rta"
complement(264..476)
/note="BLASTN similarity to 257823 (1..213): match: 1, score: 1.3e-79; database searched: nt; H.sapiens CPG island DNA genomic MseI fragment, clone 198b12, forward read cp9198b12.frla"
join(<338..421,5631..5672,10312..10374,10466..10687,12641..12796,13018..13268,15326..15467,16915..17065,17502..17578,18383..18487)
/note="Hypothetical groucho-related gene product; Hypothetical protein exhibits 52% identity to Mus musculus groucho-related gene product (AF145957) over 298 residues, and 36% identity to Tlcz_HUMAN (Q04725) over 292 residues. CDS constructed from combination of Xgral1 predictions and EST alignments"
/codon_start=1
/evidence=not_experimental
/product="R26510.1"
/protein_id="AA038076.1"
/db_xref="PID:g5030439"
/db_xref="GI:5030439"
/translation="AVFTGLEQKFGKDCIANGNDHGGPLKIQDVAENHKQIG
NVLQIVESCQLOFQSEEVPAEPSPGTPQOQKDTLDESSDITATRSSMRLR
PLGDNQPELOFMDKPEFWMHDTLEOIMFEGMILHAGELVLTATISSPFRHF
TCGRGCIKVSILTGOVADRPESHLPICPGATLPICLSSNRSLITGYNLASIS
VMDIAPSLHVKEDLPACALNCQALNDANLNANLAFSTGVSINWLRQOSVVDLK
GYDPGVKSIVKGNITWGDPDACLRCWDQRTINKLEFYGFQSQIMSLSPQDWDVL
LGNANGQOMLQSTSGQRHMVGCQSDVILSVKFSFQCMNAVGMDFLGYSPAGT
KVEFVPEMSPTCDVSSNNRLVYVTSGEHASYQITY"
1260..1858
/note="BLASTN similarity to A049067.1 (1..599): match:

0.98, score: 2.5e-237; database searched: month.nt; RPECI-11-268021.T0 RPECI-11 Homo sapiens genomic clone RPECI-11-268021, genomic survey sequence"
1951..2245
repeat_region
/rpt_family="AluLo"
complement(2442..2734)
/rpt_family="AluSg"
complement(2753..3027)
/rpt_family="AluSg"
complement(3030..3050)
/rpt_family="AT_rich"
3142..3463
repeat_region
/rpt_family="AluSg"
complement(3464..3496)
/rpt_family="MADB1"
3937..4240
repeat_region
/rpt_family="AluSx"
4252..4553
/rpt_family="AluLo"
4537..4849
repeat_region
/rpt_family="AluSg"
5173..5462
/rpt_family="AluSx"
complement(5524..5581)
/rpt_family="MIR"
5739..6038
repeat_region
/rpt_family="AluSx"
complement(6041..6181)
/rpt_family="LINE2"
complement(6684..7318)
/rpt_family="LINE2"
complement(7953..8250)
/rpt_family="AluSx"
8452..8746
repeat_region
/rpt_family="AluSx"
complement(8884..9201)
/rpt_family="AluY"
9277..9574
repeat_region
/rpt_family="AluSx"
9614..9905
/rpt_family="AluDb"
9915..10206
repeat_region
/rpt_family="AluSx"
11798..12087
/rpt_family="AluY"
13479..13595
/rpt_family="MIR"
13600..13696
/rpt_family="AluLo/FLAM"
complement(13708..13803)
/rpt_family="AT_rich"
13804..14104
/rpt_family="AluY"
14105..14194
/rpt_family="(TA)n"
14204..14486
repeat_region
/rpt_family="AluSx"
14490..14564
/rpt_family="(CA/TA)n"
14565..14855
repeat_region
/rpt_family="AluSx"
complement(14856..14959)
/rpt_family="(TA)n"
15537..15673
repeat_region
/rpt_family="AluSg/x"
15675..15964
/rpt_family="AluSp"
15972..16276
/rpt_family="AluDb"
16390..16701
repeat_region
/rpt_family="AluSx"
complement(16737..16812)
/rpt_family="MIR"

CDS

NGVDFIDISTHOYPTDIAFSTSSNMEAMAKARGELAEVKKALBEAYLPVYTEM
NNSPRDRPDPKIPYDAAFVTKTITDIDIPGCSYMTFTDIFEEBGGSLPFGHF
GLTNINGIKPSPTRAPQIOLDKNGERIEFEDKSPITDICIAYQNESEITLISNHY
PLSPIDTENIKVVLKGIENCRREVSEIDENANPNKRVIMKSPALNREQIEELIK
ASLKKREKVSQWGVNNNEITFDLSVPHSVAVATIKNG"
5439..>6067
/note="ORF 6"
/codon_start=1
/transl_table=11
/protein_id="AA23064.1"
/db_xref="PID:g144301"
/db_xref="GI:144301"
/translation="MGILYTKRANPYTEYLAKRISTLLIMTIVLELIMKPSYKK
AYAVVLANQIVGVKDKTEKQNLITQKKVEERHNDSTLOSLOLKSIEGQIRE
TRYDELKNTIIEGKVLKRYAIFNSKPYFENPQTNILNKLKRYINDKASQA
KLEKYEIKFVYVSPAIKVADEATALKIMEGDQVLEYTVKSGDTLMDP"

BASE COUNT 2230 a 787 c 1243 g 1807 t
ORIGIN 1 bp upstream of BamHI site.

Query Match 73.08; Score 16.8; DB 1; Length 6067;
Best Local Similarity 90.08; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCAGGCC 20
|||||
Db 2561 AATTACACACTGGCATTCC 2542

RESULT 13
LOCUS AF005383/c DNA BCT 27-NOV-1997
DEFINITION Caldicellulosiruptor saccharolyticus putative transport protein
(XynG), putative transport protein (XynH), xylanase (XynF),
(XynE), xylanase (XynB), xylanase (XynA), acetylxylosidase
(XynC) and xylanase (XynB) genes, complete cds.

ACCESSION AF005383
NID 92645411
VERSION AF005383.1 GI:2645411

KEYWORDS

Caldicellulosiruptor saccharolyticus.
Caldicellulosiruptor saccharolyticus
Bacteria: Firmicutes: Bacillus/Clostridium group;
Thermococcus group; Caldicellulosirupter.

AUTHORS Luthi, E., Love, D.R., McNulty, J., Wallace, C., Caughey, P.A., Saul, D.
and Bergquist, P.L.
TITLE Cloning, sequence analysis, and expression of genes encoding
xylan-degrading enzymes from the thermophile 'Caldicellu-

JOURNAL Appl. Environ. Microbiol. 56 (4), 1017-1024 (1990)
MEDLINE 90253140
REFERENCE 2 (bases 1 to 18483)
AUTHORS Te'o, V.S. Jr., Gibbs, M.D., Saul, D.J. and Bergquist, P.L.
TITLE A cluster of genes involved in xylan degradation cloned from the
extreme thermophile Caldicellulosiruptor saccharolyticus
Appl. Environ. Microbiol. (1997) In press

JOURNAL Te'o, V.S. Jr., Gibbs, M.D., Saul, D.J. and Bergquist, P.L.
REFERENCE 3 (bases 1 to 18483)
AUTHORS Direct Submission
TITLE Submitted (26-MAY-1997) School of Biological Sciences, Macquarie
JOURNAL University, Balacakra Road, Sydney, NSW 2109, Australia
FEATURES
Source Location/Qualifiers
1. 18483
/organism="Caldicellulosiruptor saccharolyticus"
/db_xref="taxon:44001"
CDS <1..937
/note="ORF1"
/codon_start=2
/transl_table=11
/protein_id="AA87367.1"
/db_xref="PID:g2645412"
/db_xref="GI:2645412"

gene

CDS

/translation="DPEGVYLGSGNFILOKMNIFESONKTKYVPDVFETFKSGSLFYPI
NVATISMARVPRKLYRSPATCYPPDDIPVFLPMLPNYGRGRVKAISPKLFTKNI
SISISIKMLSDGETQINLQTEKLEFQIVGVIKRRHYLITLDENGVITVASEIV
DIPISKFAIEDYEPFOMVLRGYIPSTSPVDMVSSMLKLVKVEGSIKKTW
TANEDBELKTEICDYLSNRYKLMIDLKRVYELKPVSKIKKKFLERHLLKGLV
LVKFEGETEILIVYTSIDQVNTVYVAKRNFYIE"
953..1873
/gene="XynG"
953..1873
/note="similar to Agrobacterium radiobacter lacG (part of
the binding-protein-dependent transport system for
lactose). Swiss-Prot Accession Number P29823"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AA87368.1"
/db_xref="PID:g2645413"
/db_xref="GI:2645413"

gene

CDS

/translation="MIRNPFQIISRKSLKREKAMYGRLFVLPVIGLIVFIPIE
VNSVYTFQKLSGDGLSEFSGWENYIYAFKDPYVKNLTSITSLDEYPIYIV
FSIFVAVILKDEFGRTFAITIFFPVYIASGVYIVLKENVAGGSSNMVASTIF
KAENLAVILNACIPRMSKXVVDIYKIPDLTRSGVOILLIALLHNIPKSFYEVA
VIEGATEMEKREKITEPMISPYVVAIYISIDYFDYGNQVNRVVDKANKRFEFS
TTAIIYELVYMIILLVNRVYIKRYYVL"
1918..2937
/gene="XynH"
1918..2937
/note="similar to Agrobacterium radiobacter lacG (part of
the binding-protein-dependent transport system for
lactose) encoded by the sequence presented in Genbank
Accession Number X66596"
/codon_start=1
/transl_table=11
/product="putative transport protein"
/protein_id="AA87369.1"
/db_xref="PID:g2645414"
/db_xref="GI:2645414"

CDS

/translation="MRNVKSLSMSYIGKRLNVEYTSVAKRRLFVLRKVPYAVY
CMYVILYPIFLIFSNALRSKEDLPSPVMIWKHITFSNFEYANVLSYPAIKNTL
LMILPVIYIOTFISMWVGFAFRERELITLITLFTVLPVLOTIVLPVYKPFY
DFYIGKLGFITGKPLVYMLDTPMPFTILNFGKIRSLITIFPPOFRMPPEL
EBAKIDGCPSTFTIRINVPNATGATITWMLFSIVHWNDIYVSATFENLPIISM
LVLANRMSMIODAYNFSSEDIYLMGSLVLEACLMVILPLIYIYIGORYTESIER
TGIVG"
3016..4800
/note="ORF2"
/codon_start=1
/transl_table=11
/protein_id="AA87370.1"
/db_xref="PID:g2645415"
/db_xref="GI:2645415"
/translation="MRKSKFEKIVSIVLVCILTSVAVYISMTQKASSTLGDIS
ELRPGSKESLKTDLFNKRVTAIEDYOKKRGKRVYISDMNNONKILAMAGD
PIDVITGSGTSPSHFTRGVLPDLKDYVDLAKAYINRPAVDYFKTNGHYLSQGS
NPMVLIYNDMLMEGIDEEMLALYKRGKNNWTFALAKRLADTKDKIDRY
GVNFWATIALYVNGVAFVSIKTKGKLEFDNAALQPALNFKKAKESMLARDMDI
TVSGLKROGVMLVAPQYKFDOKREVEDLEAPLDGDNKTKGKAPPADQGIK
GSKNPVGAKEFTNLLESYOKNHDVYNAKRPYIVLVDPYKLAOKRFPYPLGGGLMG
PHNDITGRNDSVSVALSSLRQIEKNIKNDQOTGREVYVYKQFPFTINFDGKS
FKLLDTSKRTVRLSVSGDAIKGSLKVTWDSKSGKEIYVTDPAKVIYGMHDK
ISEFVYKAVTAGKTTVSCITLISEPSSGKSGNISKITIDGQTYVYKNGNTINID
NSOFRMSRLIGITGCGFVIDNIRYEIE"
4881..8924
/gene="XynF"
4881..8924
/note="XynF"
/function="has both arabinofuranosidase and exoxylanase
activities"
/codon_start=1
/transl_table=11
/product="xylanase"

```

/gene="xynA"
/function="endoxylanase activity"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87374.1"
/db_xref="GI:2645419"
/db_xref="GI:2645416"
/translation="MANKLRLNKKILVLPALIVLCILSAFNFQINKAYSSNOOG
SKLQNNITSSNPDIISDIPDDIIRVGDVNSPTMHLTPGVPMHSTDIWNRII
GVLVDLENDNDAHNLKNGNNITGKGMACIRYKKGKFEVLGALDTGTLYSSNII
TGPWEVEFENEYLDHPSLLPDNGKAYIIYGGSEIPEKELMDIKRINKGLNKYIK
SKIDGLEESHAUKINGKYYITTTIREKGGIRECYTRDKIDGPEGVYKIDILGT
KINGVAQGLIDTPDGKMYAMLQDHDAVRPVLVPEWEGMPYEGDEKQVPLKE
VPGQSSSETEVYKSDDEFQKAEASNPREFEVRKRNALIKDELVINNPPGTLGK
GEGAKLEVVSEGNKNIHAYNRTGRSGIQONLGRVKKRKAFFVKKVYKGGD
KEFILAARTSNGKESYOEILKGFANKNEMTSISGTFISODIEYVYLIOYPIREK
DKNDLVDFEVEVASYKEVLYTOKSEEAAPGVSNTLMNNHNDPKSKMLYKRG
FLIKTCNVSDIOQARNTLTORTLGRKSGHILNDYRMKGGDVALGLAEPKEGFI
GVKQKDEYIVAVEKGEVEYTKLQOSVYVLIKDEFEIDKAVFYESTNGINWIKFG
SOLTKMYTIPHEGYNRAFAFNATKQTGGYVIDEFERKELGTEDEDFVYLLKED
IELRLNDIKRNEELVSLNSLRKKATQINILKFPRLDVEFALVADKRIKNAVVL
VEKSGDQALINKINIDTEVLDGNNMIDLVNIIKLRKREKLESTDEIRIOSLNEY
EKEKINRESAATKVFVAPKTPVCKIPPNANPLIAHKEGADPAVLAKKNNVYIYT
NDILEYDNGNKNDYTSKINKITIISSDDLNTMDHGEIEVAGRPVLYKKNATQSNAP
STAYKIKNGDKFFLIFANNASGIGVLTSDTPGVPEDEIGRPLISRTIPVEGVVVL
FDPAYLVDDGKAYIYFGGVPOGODAMPNTARVQAGDMSVGSAAVITIPAPME
DSINKIGNTYYCYCTNFFSGARPQSGPAGVIAVMTSKSPGMEYKGVILKPN
FEGVGNHGHOFEEFNKMYIAVHAQTLAKDGVAKGYRSPHITOYEIENGKIKVEIA
DVRGIAOVYFDPYKMYEAFETPACAGISTKRAMSNMCLTGIDSGDMLATSKYDEG
NYGPKFTEAMVSNINGKYEIKIDSVDKIIIVYEVQDSSTHMYAVETNVDKVT
GVHDYFVFKGSEKRLFDMCKMFKF"
/gene="xynB"
/gene="xynC"
/function="exoxylanase activity only"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87372.1"
/db_xref="GI:2645417"
/db_xref="GI:2645417"
/translation="MKRTFKSVTLIAVAIVLVSGVFLKAPVWQNAKQSSSTVTI
NEGKDTLFFAYGNKILATDQSSALIEGKSKIVNKSINWSLIDVADVLRKRTW
VISSYIKHVKRPIAFESITAYLDGKGLYVLDGEIVPDSKEMIAVWVKTLKNT
NLIIAHPVDTKTAIVNDSIOVMEETLSTKAIYKDFENNTNMRRGEGVAKIL
DSKRFEGNESLYSGRAFMHGAKIPIKYVVPVNGRFSIMVHTSDLRFSILV
ORKMADEQRYRDMITSKEVAGDGEESISYELIELEIELEIELEIELEIELEIE
DPTISDTKILQOPNTSLPSLEKRYKNDKVAAGYGLINISDKQIKKRFNSITP
GEMKPESLRGPDKYDFTIADAFVEFAIKNNISIRGHTLVHNOTPMFEKDSNGE
LKKDLRLKRLKHIYVGRYKGIYAMDVNEAIDETDPDGRSNMVKICGPYIE
KAFIWAHEADPOAKLEFNDYNTVEPOKRMFIYMKIMKMSKSIPIHGLIOCHINVS
PSVEIEIETIKLESTIPGLEIOTELDMSFYOMGSSVYVVERELMLROAKYELF
NLFKRYKTIKISVTFMGLKDDYSMLRGVNNKDPFLDEYIDGRAPAFMALDYVIP
ONINLPAPPAIKLMTKK"
/gene="xynD"
/gene="xynE"
/gene="xynF"
/function="xylosidase activity"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87373.1"
/db_xref="GI:2645418"
/db_xref="GI:2645418"
/translation="MGBGLVTNRIKIEKQIGIFPDKKKFCVSGRIGLALOKETIDA
LTVKHKHIDFKIRAHGLHDVGIREDIDGNEVSFNFYIDRIIDSELEDIRP
FEVIEFMPKSLASGQVYFWRGNVTPADYGNKWRLLKSVKHEIDRGEVEVOMP
FEINNEPNLNFMKDAQAEYFKLEYAKAKEVNENIKVGPALICGSDYMDIDL
HCYKKNYVDELTRHAATGKRPQYTPYQDVDPDIEMLEKFSYRKYVSEADP
LPIHTEVSSVHPLCPIHDPDPNAAYLAVRYSEAGDVPDSEYVVERELMLROAKYELF
AIFHGFGIYANNNTPKPYFHFTFPNNAMKILYRDEHILITEENGLIAYANNEY
MKKEATDKRIKIEIPVDINDVFIKOKLIDEEENGMWRTWIMG"
/gene="xynA"
/gene="xynB"
/gene="xynC"
/gene="xynD"
/gene="xynE"
/gene="xynF"

```

```

/gene="xynA"
/function="endoxylanase activity"
/codon_start=1
/transl_table=11
/product="xylanase"
/protein_id="AAB87374.1"
/db_xref="GI:2645419"
/db_xref="GI:2645419"
/translation="MRCITVCENEMLEMLSLAKTYKDYFKIGAAYTANDEGVHRIIL
LKHFNLSLPENAMKEFNHPEDORYNEEVARIKEFAIKNDMLGHTFVHNDTPGM
VFLDNGEASKEVIERLREHITKICRYKOVVYAMDVNEAEDKTEKLLRESNMR
KIIIGDYIKIAFEIAREVAGDAKLFYNDNNMMPKLEKTYKVLKELLEROTPIDIGIS

```

Query Match
Best local similarity 73.0%; Score 16.8; DB 2; Length 18483;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Db 14977 AATTACACACTGGCATGCC 14958

Query 1 AATTACACACTGGCATGCC 20

RESULT 14
184210/c
LOCUS 184210 2062 bp DNA
DEFINITION Sequence 8 from patent US 5695766.
ACCESSION 184210
MID 93021730
VERSION 184210.1 GI:3021730
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2062)
AUTHORS Paul, P.S., Halbur, P.G., Meng, X., Lyoo, Y.S. and Lum, M. Anne.
TITLE Highly virulent porcine reproductive and respiratory syndrome
viruses which produce lesions in pigs and vaccines that protect
pigs against said syndrome
Patent: US 5695766-A 8 09-DEC-1997;
FEATURES
source location/Qualifiers
BASE COUNT 480 a 491 c 530 g 561 t
ORIGIN
Query Match
Best local similarity 73.0%; Score 16.8; DB 5; Length 2062;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTACACACTGGCATGCC 20

Db 1935 AATTCAACACTGGCATGCC 1916

RESULT 15
184217/c
LOCUS 184217 164 bp DNA
DEFINITION Sequence 19 from patent US 5695766.
ACCESSION 184217
MID 93021737
VERSION 184217.1 GI:3021737
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 164)
AUTHORS Paul, P.S., Halbur, P.G., Meng, X., Lyoo, Y.S. and Lum, M. Anne.
TITLE Highly virulent porcine reproductive and respiratory syndrome
viruses which produce lesions in pigs and vaccines that protect
pigs against said syndrome
Patent: US 5695766-A 19 09-DEC-1997;
JOURNAL
FEATURES location/Qualifiers

source 1.164
BASE COUNT 50 a /organism="unknown"
ORIGIN 28 c 45 g 41 t

Query Match 73.0% Score 16.8; DB 5; Length 164;
Best Local Similarity 90.0% Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AATTCACACACTGGCATGCC 20
||||| ||||| |||||
Db 37 AATTCACACACTGGCATGCC 18

Search completed: September 13, 1999, 15:55:19
Run time: 4547 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:29 ; Search time 255.05 Seconds

(without alignments)
22.562 Million cell updates/sec

Title: US-09-325-095-15

Perfect score: 23

Sequence: 1 AATTGACACACTGGCATGCCGAT 23

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	23	100.0	3207	1	051155	p110 CDNA. Recombi
C 2	23	100.0	23	1	059014	RACE PCR primer 22
C 3	23	100.0	3498	1	057012	Pitins 3-kinase 11
C 4	23	100.0	23	1	057012	Pitins 3-kinase pr
C 5	21.4	93.0	3412	1	051156	Human p110 CDNA. R
C 6	16.8	73.0	1938	1	063591	ISF-12 3' terminal
C 7	16.8	73.0	164	1	063598	ISF-12 non-transla
C 8	16.8	73.0	164	1	063598	PRSV VR 2385 3'-n
C 9	16.8	73.0	2062	1	051389	PRSV VR 2385 3' t
C 10	16.6	72.2	5559	1	055260	Restriction fragme
C 11	16.6	72.2	5559	1	052600	E11 genomic DNA fr
C 12	16.6	72.2	5559	1	052600	Pseudomonas fluore
C 13	16.6	72.2	5559	1	052600	5.6 kb EcoRI-Hind
C 14	16.6	72.2	5559	1	052600	Pseudomonas fluore
C 15	16.6	72.2	5559	1	052600	Staphylococcus aur
C 16	16.6	72.2	5559	1	052600	Staphylococcus aur
C 17	15.4	67.0	2680	1	064579	Myc-binding zinc-f
C 18	15.4	67.0	150	1	064579	Human biallelic po
C 19	15.4	67.0	203	1	064579	Human secreted pro
C 20	15.2	66.1	2375	1	032001	Encodes elmeria an
C 21	15.2	66.1	4374	1	072914	Murine receptor ty
C 22	15.2	66.1	3358	1	072914	VR-2332 Viral sequ
C 23	15.2	66.1	298	1	072914	Human gene signatu
C 24	15.2	66.1	2375	1	072914	Staphylococcus aur
C 25	15.2	66.1	2375	1	072914	Staphylococcus aur
C 26	15.2	66.1	2375	1	072914	Ethylene receptor
C 27	15.2	66.1	2375	1	072914	Sequence of brain
C 28	15.2	66.1	2375	1	072914	p110 CDNA. Recombi
C 29	15.2	66.1	2375	1	072914	Pitins 3-kinase 11
C 30	15.2	66.1	2375	1	072914	Human cytoplasmic
C 31	15.2	66.1	2375	1	072914	MEKK4 CDNA. New sl
C 32	15.2	66.1	2375	1	072914	H. pylori transmem
C 33	15.2	66.1	2375	1	072914	H. pylori surface
C 34	15.2	66.1	2375	1	072914	H. pylori ORF hp3p
C 35	15.2	66.1	2375	1	072914	H. pylori cellular
C 36	15.2	66.1	2375	1	072914	Probe HP-60 used t
C 37	15.2	66.1	2375	1	072914	CDNA encoding a mi
C 38	15.2	66.1	2375	1	072914	DNA encoding prote
C 39	15.2	66.1	2375	1	072914	MEKK4 protein codi
C 40	15.2	66.1	2375	1	072914	Enterococcus faeca
C 41	15.2	66.1	2375	1	072914	Human kidney amino
C 42	14.8	64.3	2297	1	078986	Cobra venom mocaeh
C 43	14.8	64.3	2288	1	078986	Cobra venom mocaeh

ALIGNMENTS

44	14.8	64.3	2309	1	V07899	Cobra venom mocaeh
45	14.8	64.3	1820	1	V07900	Cobra venom mocaeh
ALIGNMENTS						
RESULT 1						
ID	051155/c					
AC	051155 standard; CDNA; 3207 BP.					
DE	051155:					
DE	12-APR-1994 (first entry)					
DE	p110 CDNA.					
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;					
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;					
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;					
FW	ss.					
FW	Key	Location/Qualifiers				
FT	cds	1..3207				
FT	/*tag- a	/note= "PI3- kinase p110"				
FW	MO9321328-A.					
PD	28-OCT-1993.					
PR	13-APR-1993; G00761.					
PR	13-APR-1992; GB-008135.					
PA	(LUDW-) LUDWIG INST CANCER RES.					
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;					
PI	Parkerp], Volinia S, Waterfield MD;					
DR	WPI: 93-351738/44.					
DR	P-RSD: R43341.					
PT	Recombinant polypeptide(s) - with phosphoinositide-3 kinase					
PT	activity, useful for controlling cell proliferation					
PS	Claim 7; Fig 9; 146pp; English.					
CC	Southern blot analysis was performed using a bovine CDNA probe contg.					
CC	a fragment of a PI3-kinase-encoding sequence and human CDNA isolated					
CC	from a CDNA library constructed from mRNA isolated from the human					
CC	cell line K562. Positive clones were sequenced to give the human					
CC	PI3 kinase p110 sequence shown. This sequence has 95 percent					
CC	homology with the bovine sequence. The domain encoding residues 19-					
CC	100 of human p110 is sufficient to encode the kinase which will					
CC	associate with the p85 kinase subunit. The gene may be used to					
CC	provide a protein with PI3 kinase activity, and is useful for					
CC	screening for (ant)agonists of PI3 kinase activity which could be					
CC	useful for stimulation or inhibition of cell proliferation and hence					
CC	prophylaxis or therapy. Platelet or neutrophil activity or blood					
CC	glucose levels can be controlled using the kinase.					
CC	See also 051155, 059012-23 and 057522-3.					
SQ	Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;					
Query Match 100.0%; Score 23; DB 1; Length 3207;						
Best Local Similarity 100.0%; Pred. No. 0.014;						
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;						
OY	1 AATTGACACACTGGCATGCCGAT 23					
DB	383 AATTGACACACTGGCATGCCGAT 361					
RESULT 2						
ID	059014					
AC	059014 standard; DNA; 23 BP.					
DE	059014:					
DE	12-APR-1994 (first entry)					
DE	RACE PCR primer 2224.					
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;					
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;					
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;					
OS	ss; amplification; p110.					
PN	MO9321328-A.					
PD	28-OCT-1993.					
PF	13-APR-1993; G00761.					

PR 13-APR-1993: GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parkerp], Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 39; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC as primers. Products were fractionated using an agarose gel. The
 CC cDNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also 051155-6, 059012-23 and 057522-3.
 SO Sequence 23 BP; 7 A; 7 C; 4 G; 5 T;

Query Match 100.0%; Score 23; DB 1; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.0065;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCGAT 23
 Db 1 AATTCACACACTGGCATGCCGAT 23

RESULT 3

057012/c
 ID 057012 standard; cDNA to mRNA; 3498 BP.

AC 057012;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key
 FT 1.3207
 FT /tag- a
 FT /product- p110

PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1993; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PsDB: R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure: Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylinositol (Ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the Ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match

100.0%; Score 23; DB 1; Length 3498;

Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCGAT 23
 Db 383 AATTCACACACTGGCATGCCGAT 361

RESULT 4

057020
 ID 057020 standard; DNA; 23 BP.

AC 057020;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase primer oligo 2224.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1993; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 15; 71pp; English.

CC The sequences given in 057020-23 are primers which were used in the
 CC isolation of the phosphatidylinositol (Ptdins) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schizosaccharomyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the Ptdins is induced. Ptdins activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SO Sequence 23 BP; 7 A; 7 C; 4 G; 5 T;

Query Match 100.0%; Score 23; DB 1; Length 23;
 Best Local Similarity 100.0%; Pred. No. 0.0065;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCGAT 23
 Db 1 AATTCACACACTGGCATGCCGAT 23

RESULT 5

051156/c
 ID 051156 standard; cDNA; 3412 BP.

AC 051156;
 DT 12-APR-1994 (first entry)
 DE Human p110 cDNA.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ds.
 OS Human.
 FH Key
 FT 1.3207
 FT /tag- a
 FT /note- "P13- kinase p110"
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.

```

PR 13-APR-1992: GB-008135.
PA (LUDM-) LUDMIG INST CANCER RES.
PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parkerj, Volinia S, Waterfield MD;
DR WPI: 93-351738/44.
DR P-PSDB: R43342.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146pp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line KG1a. Positive clones were sequenced to give the human
CC p13 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with p13 kinase activity, and is useful for
CC screening for (ant)agonists of p13 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also Q51155 and Q57522-3.
CC Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match
Best Local Similarity 93.0%; Score 21.4; DB 1; Length 3412;
Matches 22; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCCAT 23
DB 383 AATTCACACACTGGCATGCCAT 361

RESULT 6
063591/C
AC 063591 standard; cDNA; 1938 BP.
DT 12-DEC-1994 (first entry)
DE ISU-12 3' terminal sequence.
KW Primer: polymerase chain reaction; PCR; amplify; probe: Iowa strain;
KW infectious agent; porcine respiratory and reproductive syndrome;
KW ISU-12; vaccine; porcine respiratory and reproductive disease; PRRD;
KW antibody; assay; ss.
OS Porcine respiratory and reproductive disease virus, ISU-12.
FH Key Location/Qualifiers
FT cds 239..901
FT /*tag= a
FT /label= ORF-5
FT cds 889..1403
FT /*tag= b
FT /label= ORF-6
FT cds 1403..1771
FT /*tag= c
FT /label= ORF-7
FT EP-595436-A.
PN 04-MAY-1994.
PD 29-OCT-1993; 203042.
PF 30-OCT-1992; US-969071.
PR 05-OCT-1993; US-131625.
PA (SOLV) SOLVAY ANIMAL HEALTH INC.
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
PI Halbur PG, Lum MA, Lyoo XS, Meng X, Paul PS;
DR WPI: 94-146025/18.
DR P-PSDB: R53462.
PT New porcine respiratory and reproductive disease virus - used to
PT prepare vaccines and antibodies for diagnosis, treatment and
PT prophylaxis of virus infection.
PS Example 4; Page 45-47; 98pp; English.
CC This sequence represents the 3' terminal region of the infectious
CC agent associated with the Iowa strain of porcine respiratory and
CC reproductive syndrome, termed ISU-12. Three large open reading

```

```

CC frames (ORFs) have been found in this sequence, ORF-5, ORF-6 and
CC ORF-7 (see also Q63592-94). These ORFs all have the capacity of
CC encoding more than 100 amino acids. ORF-5 and ORF-6 overlap each
CC other by 10 bp and ORF-6 and ORF-7 overlap each other by 5 bp. This
CC isolated ISU-12 sequence may be used to infect cells and from these,
CC the vaccine of the invention can be produced. This vaccine may be
CC used for protecting pigs against a porcine respiratory and reproductive
CC disease (PRRD). Antibodies to the vaccine may also be used in treating
CC PRRD and for assaying for the virus.
SQ Sequence 1938 BP; 453 A; 458 C; 497 G; 530 T;

Query Match
Best Local Similarity 73.0%; Score 16.8; DB 1; Length 1938;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCC 20
DB 1811 AATTCACACACTGGCATGCC 1792

RESULT 7
063598/C
AC 063598 standard; cDNA; 164 BP.
DT 12-DEC-1994 (first entry)
DE ISU-12 non-translated sequence.
KW Primer: polymerase chain reaction; PCR; amplify; probe: Iowa strain;
KW infectious agent; porcine respiratory and reproductive syndrome;
KW ISU-12; vaccine; porcine respiratory and reproductive disease; PRRD;
KW antibody; assay; ss.
OS Porcine respiratory and reproductive disease virus, ISU-12.
PN EP-595436-A.
PD 04-MAY-1994.
PF 29-OCT-1993; 203042.
PR 30-OCT-1992; US-969071.
PR 05-OCT-1993; US-131625.
PA (SOLV) SOLVAY ANIMAL HEALTH INC.
PA (IOWA) UNIV IOWA STATE RES FOUND INC.
PI Halbur PG, Lum MA, Lyoo XS, Meng X, Paul PS;
DR WPI: 94-146025/18.
PT New porcine respiratory and reproductive disease virus - used to
PT prepare vaccines and antibodies for diagnosis, treatment and
PT prophylaxis of virus infection.
PS Example 4; Page 53; 98pp; English.
CC This sequence represents the non-translated sequence from the 3'
CC terminal region of the infectious agent associated with the Iowa
CC strain of porcine respiratory and reproductive syndrome, termed
CC ISU-12. This sequence ends the ISU-12 fragment which contains open
CC reading frames (ORFs) ORF-5, ORF-6 and ORF-7. These ORFs all have
CC the capacity of encoding more than 100 amino acids. ORF-5 and ORF-6
CC overlap each other by 10 bp and ORF-6 and ORF-7 overlap each other by
CC 5 bp. The isolated ISU-12 3' sequence may be used to infect cells and
CC from these infected cells the vaccine of the invention can be produced.
CC This vaccine may be used for protecting pigs against a porcine
CC respiratory and reproductive disease (PRRD). Antibodies to the
CC vaccine may also be used in treating PRRD and for assaying for the
CC virus.
SQ Sequence 164 BP; 50 A; 28 C; 45 G; 41 T;

Query Match
Best Local Similarity 73.0%; Score 16.8; DB 1; Length 164;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCC 20
DB 37 AATTCACACACTGGCATGCC 18

RESULT 8
T14394/C
ID T14394 standard; cDNA; 164 BP.

```

AC T14394:
DT 05-AUG-1996 (first entry)
DE PRRSV VR 2385 3'-nontranslational region.
KW PRRSV; vaccine; antigen; ss.
OS pig reproductive and respiratory syndrome virus Iowa (VR 2385).
PN MO9606619-A1.
PD 07-MAR-1996.
PF 01-SEP-1995: U10904.
PR 01-SEP-1994: US-301435.
PA (HALB/) HALBUR P.
PA (LUMM/) LUM M A.
PA (MENG/) MENG X.
PA (MORO/) MOROZOV I.
PA (PAUL/) PAUL P S.
PI Halbur P, Lum MA, Meng X, Morozov I, Paul PS;
DR WPI: 96-160132/16.
PT New porcine reproductive and respiratory syndrome virus DNA - and
PT proteins encoded by open reading frames of an Iowa strain of the
PT virus; are used in vaccines against PRRSV in pigs
PS Disclosure: Page 147: 228pp; English.
CC The 3'-nontranslational sequence following the stop codon of
CC ORF-7 (see also T14392) of porcine reproductive and respiratory
CC syndrome virus (PRRSV) strain Iowa isolate ISU-12 (VT 2385) was
CC identified in the genome of VR 2385 and was compared to the
CC corresponding sequence (T14397) of PRRSV Ielystad, a European strain.
CC A consensus region adjacent to the polyA tail was identified.
CC homology was found between the 2 sequences.
SQ Sequence 164 BP; 50 A; 28 C; 45 G; 41 T;

Query Match 73.0%; Score 16.8; DB 1; Length 164;
Best Local Similarity 90.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCC 20
DB 37 AATTCACACACTGGCATGCC 18

RESULT 9
T14389/C
ID T14389 standard; cDNA; 2062 BP.
AC T14389;
DT 04-AUG-1996 (first entry)
DE PRRSV VR 2385 3' terminal sequence.
KW PRRSV; vaccine; antigen; ss.
OS pig reproductive and respiratory syndrome virus Iowa strain ISU-12.
FH Key location/Qualifiers
FT cds 426..1028
FT /*tag- a
FT /label- ORF-5
FT 1013..1037
FT /*tag- b
FT /label- ORF-6
FT /product- viral membrane protein
FT 1527..1898
FT /*tag- c
FT /label- ORF-7
FT /product- capsid protein
PN MO9606619-A1.
PD 07-MAR-1996.
PF 01-SEP-1995: U10904.
PR 01-SEP-1994: US-301435.
PA (HALB/) HALBUR P.
PA (LUMM/) LUM M A.
PA (MENG/) MENG X.
PA (MORO/) MOROZOV I.
PA (PAUL/) PAUL P S.
PI Halbur P, Lum MA, Meng X, Morozov I, Paul PS;
DR WPI: 96-160132/16.
DR P-PSDB: R94701;
DR P-PSDB: R94702;
DR P-PSDB: R94703.

PT New porcine reproductive and respiratory syndrome virus DNA - and
PT proteins encoded by open reading frames of an Iowa strain of the
PT virus; are used in vaccines against PRRSV in pigs
PS Disclosure: Page 139-140: 228pp; English.
CC The 3' terminal sequence (T14389) of a porcine reproductive and
CC respiratory syndrome virus (PRRSV) Iowa strain isolate ISU-12 (VR
CC 2385) cDNA clone was detd. following PCR amplification (see also
CC T14377-83). The sequence includes ORFs 5 (see also T14390), 6
CC (T14391) and 7 (T14391), whose products (R94701-03) can be used in
CC the development of pig PRRSV vaccines.
SQ Sequence 2062 BP; 480 A; 491 C; 530 G; 561 T;

Query Match 73.0%; Score 16.8; DB 1; Length 2062;
Best Local Similarity 90.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AATTCACACACTGGCATGCC 20
DB 1935 AATTCACACACTGGCATGCC 1916

RESULT 10
ID 055260
AC 055260;
DT 28-JUL-1994 (first entry)
DE Restriction fragment comprising gene activating sequence.
KW Gene activation; gafa; transcriptional activator; activator;
KW sensor; regulation; induction; biocontrol agents; crop protection;
KW ss.

OS Pseudomonas fluorescens.
FH Key location/Qualifiers
FT cds 210..1688
FT /*tag- a
FT /label- ORF 1.
FT 1906..3633
FT /*tag- b
FT /label- ORF 2.
FT 4616..4691
FT /*tag- c
FT /label- g1yw
FT /note- tRNA coding sequence.
FT cds 4731..5318
FT /*tag- d
FT /label- ORF 3.

PN MO9401561-A.
PD 20-JAN-1994.
PF 02-JUL-1993: U06300.
PR 02-JUL-1992: US-908284.
PA (CIBA) CIBA GEIGY AG.
PI Becker J, Garfney TD, Hill DS, Howell CR, Lam ST;
PI Ilyon JM, Stein JT;
DR WPI: 94-035067/04.
PT Gene activating sequences, part. from Pseudomonas - used for
PT inducing expression of genes which are latent or expressed at low
PT levels in target organisms
PS Disclosure: Page 71-74; 87pp; English.
CC The gene activating sequence is contained within an 11 kilobase
CC EcoRI fragment (E11) from P. fluorescens. DNA sequence analysis of
CC E11 has identified five open reading frames (ORF), of which ORFs
CC and ORF 2 share significant homology with numerous activator and
CC sensor components respectively of bacterial two-component
CC regulatory systems. The gene activating sequences can be used in
CC microorganisms to induce the production of gene products and
CC secondary metabolites. They are used particularly for production
CC of bacterial biocontrol agents for use in inhibiting plant
CC pathogens. This sequence is a fragment of E11. See also 055261.
SQ Sequence 5539 BP; 1194 A; 1743 C; 1637 G; 984 T;

Query Match 72.2%; Score 16.6; DB 1; Length 5559;
Best Local Similarity 82.6%; Pred. No. 23;

[illegible]

```

FT      /*tag- e
FT      /product- "uvr exonuclease subunit C"
FT      /note- "coding sequence for uvr exonuclease subunit C
FT      has homology to uvrC"
FT      complement (7400..8041)
FT      CDS
FT      /*tag- f
FT      /function- "response regulator/transcription activator"
FT      /product- "gacA (aka gafa)"
FT      /note- "coding sequence for gacA (aka gafa) has homology
FT      to the uvrC and gacA genes of E.coli and
FT      Ps.fluorescens, respectively"
FT
FT      W09824919-A1.
FT      11-JUN-1998.
FT      05-DEC-1997; E06815
FT      09-SEP-1997; US-058304.
FT      06-DEC-1996; US-761258.
FT      (NOVS ) NOVARTIS AG.
FT      Gaffney TD, Hill DS, Lam ST, Ligon JM, Stafford JM,
FT      Torkewitz NR:
FT      WPI: 98-33337/29.
FT      Genetically modified Pseudomonas strains - useful to protect crop
FT      plants by controlling or inhibiting plant pathogen growth, e.g.
FT      growth of Rhizoctonia species
FT      PS
FT      Example 1: Page 36-45; 85pp; English.
FT      A genetically engineered biocontrol strain of Pseudomonas has been
FT      developed that can control attacks on crop plants by pathogenic fungi,
FT      e.g. Rhizoctonia and Pythium and aggressively compete with indigenous
FT      bacteria and microflora in the plant rhizosphere. The strains can be
FT      included with agronomically acceptable carriers or chemical fungicides
FT      (e.g. metalaxyl compounds) in biocontrol compositions. The strains or
FT      compositions can be applied to a plant/plant part to protect it from a
FT      plant pathogenic fungus, by controlling or inhibiting fungal growth.
FT      They can also be applied to the environment in which a plant pathogenic
FT      fungus will grow (e.g. soil) to similarly control or inhibit pathogen
FT      growth, or to seeds to protect plants developing from the seed from a
FT      plant pathogenic fungus. They are especially effective against
FT      CC plant pathogenic fungi.
FT      CC Rhizoctonia and Pythium species which cause damping off in cotton.
FT      CC Rhizoctonia also infects many other crop species (e.g. beans and wheat),
FT      and no effective chemical fungicides are available. The present sequence
FT      CC represents the 11 kb EcoRI fragment containing the gacA gene derived
FT      CC from the chromosome of Pseudomonas fluorescens strain CGA267356, used in
FT      CC an example from the present invention.
FT      SQ
FT      Sequence 10763 BP: 2197 A; 3450 C; 3136 G; 1980 T;

```

```

Query Match 72.2%; Score 16.6; DB 1; Length 10763;
Best Local Similarity 82.6%; Pred. No. 26;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 AATTCACACTGCATGCCGAT 23
DB      3394 ATTTCAGACATCGCATGCCGAT 3416

```

```

Search completed: September 13, 1999, 15:59:30
Job time: 4797 sec

```

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:55 ; Search time 1694.61 Seconds

(without alignments)
26.772 Million cell updates/sec

Title: US-09-325-095-15

Perfect score: 23
Sequence: 1 AATTCACACACTGGCATGCCGAT 23

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	23	100.0	252	32	AA331546	AA331546 EST35390
C 2	17.8	77.4	478	43	AI102374	AI102374 EST211663
C 3	17.8	77.4	535	45	AA874982	AA874982 UT-R-EO-C
C 4	17.8	77.4	358	45	AI009082	AI009082 EST203533
C 5	17.8	77.4	497	48	AI547507	AI547507 UT-R-C3-S
C 6	17.8	77.4	436	48	AI576472	AI576472 UT-R-G0-U
C 7	17.2	74.8	376	30	AA196181	AA196181 zP96a06.s
C 8	17.2	74.8	316	31	AA325642	AA325642 EST28660
C 9	16.8	73.0	465	28	AA105882	AA105882 m184h07.r
C 10	16.8	73.0	465	29	AA164056	AA164056 m167g06.r
C 11	16.8	73.0	376	30	AA272905	AA272905 v839d01.r
C 12	16.8	73.0	398	30	AA273650	AA273650 vB2c09.r
C 13	16.8	73.0	491	31	AA289192	AA289192 vB3c01.r
C 14	16.8	73.0	447	41	AU017898	AU017898 AU017898
C 15	16.8	73.0	91	44	AI133597	AI133597 EST051 MO
C 16	16.8	73.0	447	44	AI332915	AI332915 m167g06.y
C 17	16.8	73.0	412	44	AI332858	AI332858 m167g06.x
C 18	16.8	73.0	287	46	AI448486	AI448486 mp63f07.x
C 19	16.8	73.0	439	47	AI462305	AI462305 ub62b06.x
C 20	16.8	73.0	216	49	AV007110	AV007110 AV007110
C 21	16.8	73.0	403	49	AV008847	AV008847 AV008847
C 22	16.6	72.2	539	21	T83189	T83189 yd41g02.r1
C 23	16.6	72.2	355	22	R34570	R34570 yh85a12.s1
C 24	16.4	71.3	482	34	AA530221	AA530221 vJ38a08.r
C 25	16.4	71.3	265	36	AA638506	AA638506 v054b01.r
C 26	16.4	71.3	474	42	AU023168	AU023168 AU023168
C 27	16.4	71.3	633	48	AI601278	AI601278 fC09a05.x
C 28	16.4	71.3	191	49	AV006692	AV006692 AV006692
C 29	16.4	71.3	180	50	AV030936	AV030936 AV030936
C 30	16.2	70.4	340	25	N43682	N43682 SW3ICAI163S
C 31	16.2	70.4	651	29	AA130056	AA130056 zM62e11.r
C 32	16.2	70.4	473	39	AA880834	AA880834 vx46d07.r
C 33	16.2	70.4	555	44	AI294344	AI294344 LP07690.5
C 34	16.2	70.4	480	49	AI638891	AI638891 AEMTARK20
C 35	16.2	70.4	662	51	AI721941	AI721941 fC25d12.x
C 36	16	69.6	326	46	AI425258	AI425258 vv74d05.y
C 37	16	69.6	412	46	AI451931	AI451931 mp74f04.x
C 38	16	69.6	468	47	AI530531	AI530531 mp74f04.y
C 39	15.8	68.7	185	20	Z41551	Z41551 HSCGXK122.n
C 40	15.8	68.7	339	24	H85019	H85019 yS88e04.s1
C 41	15.8	68.7	549	28	AA121732	AA121732 zN80n05.s
C 42	15.8	68.7	508	29	AA179551	AA179551 zp49b12.s
C 43	15.8	68.7	302	33	AA409963	AA409963 EST01904
C 44	15.8	68.7	435	40	AA938671	AA938671 OC08h07.s
C 45	15.8	68.7	341	41	AI028020	AI028020 oy73b03.x

ALIGNMENTS

RESULT 1
AA331546/c
LOCUS
DEFINITION
ACCESSION
NTD

AA331546 252 bp mRNA
EST35390 Embryo, 8 week I Homo sapiens CDNA 5' end similar to
phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
AA331546
g1983787

```

VERSION      AA31546.1  GI:1983787
KEYWORDS
SOURCE       human
ORGANISM     Homo sapiens
REFERENCE    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Homiidae; Homo.
              1 (bases 1 to 252)
AUTHORS      Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
              Bull,C.J., Lee,N.H., Kirnness,E.F., Weinstein,K.G., Gockayne,J.D.,
              White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
              Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
              Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
              Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinle,P.S., Jr.,
              Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,M.,
              Moreno-Palancues,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
              Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
              Small,K.V., Spriggs,T.A., Uteback,T.R., Weidman,D.F., Li,Y.,
              Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
              Dimke,D., Feng,D.-F., Ferris,A., Fischer,C., Hastings,G.A.,
              He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
              Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
              Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
              Dillion,P.J., Fannon,M.R., Rosen,C.A., Hasseltine,W.A., Fields,C.,
              Fraser,C.M. and Venter,J.C.
              Initial assessment of human gene diversity and expression patterns
              based upon 83 million nucleotides of cDNA sequence
              Nature 377 (6547 Suppl), 3-174 (1995)
JOURNAL      96026280
MEDLINE
COMMENT      On Sep 12, 1996 this sequence version replaced gi:1404766.
              Other ESTs: THC168479
              Contact: Kerlavage, AR
              Bioinformatics
              The Institute for Genomic Research
              9712 Medical Center Drive, Rockville, MD 20850 USA
              Tel: 3018699056
              Fax: 3018699423
              Email: arkerlavage@tigr.org
              For clone availability, additional sequence and expression
              information related to this EST, please check the TIGR Human Gene
              Index (http://www.tigr.org/tdb/hgi/hgi.html)
              Seq primer: M13 Reverse.
FEATURES
  source
    1..252
    /organism="Homo sapiens"
    /db_xref="ATCC (Inhost):133163"
    /db_xref="taxon:9606"
    /clone_1lb="Embryo, 8 week 1"
    /dev_stage="Embryo, 8 wks"
    /note="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
    Site_1: EcoRI; Site_2: XhoI"
BASE COUNT   78 a 46 c 53 g 75 t
ORIGIN
Query Match      100.0%; Score 23; DB 32; Length 252;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AATTCACACACTGGCATGCCGAT 23
Db 119 AATTCACACACTGGCATGCCGAT 97
RESULT 2
LOCUS      A1102374      478 bp      mRNA      EST      31-JAN-1999
DEFINITION ESR11663 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
ACCESSION  A1102374
NID        93707163
VERSION    A1102374.1  GI:3707163
KEYWORDS   EST.
SOURCE     Rattus sp.

```

```

ORGANISM     Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE    1 (bases 1 to 478)
AUTHORS      Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
              Kerlavage,A.R. and Adams,M.D.
              Rat Genome Project: Generation of a Rat EST (RST) Catalog & Rat
              Gene Index
              Unpublished (1998)
JOURNAL      On Aug 21, 1998 this sequence version replaced.
COMMENT
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
FEATURES
  source
    1..478
    /organism="Rattus sp."
    /db_xref="ATCC (Inhost):2029673"
    /db_xref="taxon:10118"
    /clone_1lb="Normalized rat brain, Bento Soares"
    /note="Organ: brain; Vector: pRT3pac; Site_1: EcoRI;
    Site_2: NotI"
BASE COUNT    154 a 96 c 99 g 129 t
ORIGIN
Query Match      77.4%; Score 17.8; DB 43; Length 478;
Best Local Similarity 90.5%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 3 TTCACACACTGGCATGCCAT 23
Db 105 TTCACACACTGGCATGCCAT 125
RESULT 3
LOCUS      AA874982      536 bp      mRNA      EST      05-FEB-1999
DEFINITION UT-R-EO-clf-c-06-0-UT.s1 UT-R-EO Rattus norvegicus cDNA clone
              UT-R-EO-clf-c-06-0-UT 3' similar to g11669534[dbj|D67015|D67015
              Mouse mRNA for scg, complete cds, mRNA sequence.
ACCESSION    AA874982
NID          94230862
VERSION      AA874982.1  GI:4230862
KEYWORDS     EST.
SOURCE       Norway rat.
ORGANISM     Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE    1 (bases 1 to 536)
AUTHORS      Bonaldo,M.F., Lennon,G. and Soares,M.B.
              Normalization and subtraction: two approaches to facilitate gene
              discovery
              Genome Res. 6 (9), 791-806 (1996)
JOURNAL      97044477
MEDLINE
COMMENT      On Mar 20, 1998 this sequence version replaced gi:2979930.
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence tag present in the cDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 8-day-Embryo library. cDNA library Preparation: M. Fatima

```

Bonaldo, Ph.D. Clone distribution: clones will be available through Research Genetics The following repetitive elements were found in this cDNA sequence: 1-35, >AT_Rich#Low_complexity
Seq primer: M13 Forward.

FEATURES
source Location/Qualifiers

1..536
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/map="x"
/clone="UI-R-E0-cf-c-06-0-UI"
/dev_stage="embryonic"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: NotI; Site_2: EcoRI; This library consists of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture."
BASE COUNT 169 a 105 c 108 g 154 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 45; Length 536;
Best Local Similarity 90.5%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCACACACTGGCATGCCGAT 23
11111111111111111111
Db 122 TTCACAGACTGGCATGCCAAT 142

RESULT 4
LOCUS A1009082 358 bp mRNA EST 08-JAN-1999
DEFINITION EST203533 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
REMARKS 3' end, mRNA sequence.
ACCESSION A1009082
NID 94132989
VERSION A1009082.1 GI:4132989
KEYWORDS EST.
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
Lee, N.H., Glodex, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat Gene Index
Unpublished (1998)
JOURNAL On Jun 15, 1998 this sequence version replaced gi:3222914.
COMMENT

CONTACT: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13-21.
Location/Qualifiers

FEATURES
source 1..358
/organism="Rattus sp."
/db_xref="taxon:10116"
/map="x" p12-q11
/clone="REMB132"
/dev_stage="embryo 8, 12, 18 dpc"
/note="Vector: pRT3D-Pac; Site_1: EcoRI; Site_2: NotI."

BASE COUNT 97 a 77 c 69 g 115 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 45; Length 358;
Best Local Similarity 90.5%; Pred. No. 33;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TTCACACACTGGCATGCCGAT 23
11111111111111111111
Db 257 TTCACAGACTGGCATGCCAAT 237

RESULT 5
LOCUS A1547507 497 bp mRNA EST 22-MAR-1999
DEFINITION UI-R-C3-s1-g-08-0-UI-s1 UI-R-C3 Rattus norvegicus cDNA clone
UI-R-C3-s1-g-08-0-UI 3', mRNA sequence.
ACCESSION A1547507
NID 94464995
VERSION A1547507.1 GI:4464995
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 497)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187163.

CONTACT: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msquares@iuii.weeb.uiova.edu
The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward.

FEATURES
source Location/Qualifiers

1..497
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/map="9 q22-q33; 898A03; 1; 1p32.1-1p33; 19"
/clone="UI-R-C3-s1-g-08-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; The UI-R-C3 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, muscle, and 8, 12 and 18-day embryos, after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: UI-R-C3, UI-R-C2p, UI-R-C1, UI-R-C0, UI-R-A1, UI-R-E1. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C3) was constructed as follows: PCR amplified cDNA inserts from UI-R-C2p clones from which 3' ESTs had

been derived was used as a driver in a hybridization with the UI-R-C2p library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C3 library. This procedure has been previously described (Bonaldio, Lennon and Soares, Genome Research 6:791-806, 1996).

BASE COUNT 155 a 100 c 99 g 143 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 48; Length 497;
Best Local Similarity 90.5%; Pred. No. 35;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TTCACACTGGCAGCCGAT 23
||||| ||||||| ||
Db 119 TTCACACTGGCAGCCGAT 139

RESULT 6
A1576472 436 bp mRNA EST 05-APR-1999
LOCUS UI-R-GO-us-e-01-0-UI.s1 UI-R-GO Rattus norvegicus cDNA clone
DEFINITION UI-R-GO-us-e-01-0-UI 3', mRNA sequence.
ACCESSION A1576472
NID A1576472
VERSION A1576472.1 GI:4560848
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 436)
AUTHORS Bonaldio, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704447
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188268.

Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.wieg.uiowa.edu

The sequence contained an oligo-dt track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dt track served to verify it as a clone from the normalized ganglia library cDNA library preparation. M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)
Seq primer: M13 forward.
Location/Qualifiers
1. 436

FEATURES
source

/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-GO-us-e-01-0-UI"
/clone_lib="UI-R-GO"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI. The UI-R-GO library is a normalized library constructed from a mixture of rat tissues (nodose ganglia, dorsal root ganglia, and trigeminal ganglia). The tag is a string of

6 nucleotides present between the Not I site and the oligo-dt track. The library was constructed as described by Bonaldio, Lennon and Soares, Genome Research 6: 791-806, 1996."

BASE COUNT 140 a 79 c 93 g 124 t
ORIGIN

Query Match 77.4%; Score 17.8; DB 48; Length 436;
Best Local Similarity 90.5%; Pred. No. 34;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 TTCACACTGGCAGCCGAT 23
||||| ||||||| ||
Db 119 TTCACACTGGCAGCCGAT 139

RESULT 7
AA196181/c 376 bp mRNA EST 10-MAR-1998
LOCUS zp96a06.s1 Striatogene muscle 937209 Homo sapiens cDNA clone
DEFINITION IMAGE:628018 3', mRNA sequence.
ACCESSION AA196181
NID 91791763
VERSION AA196181.1 GI:1791763
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 376)
AUTHORS Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Krizman, D., Kucada, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Scheinberg, K., Stepien, M., Tan, F., Thaising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
TITLE WashU-NCI human EST Project
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397572.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 118 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 213.
Location/Qualifiers
1. 376

FEATURES
source

/organism="Homo sapiens"
/db_xref="GDB:5048252"
/db_xref="taxon:9606"
/map="17q21"
/clone="IMAGE:628018"
/clone_lib="Striatogene muscle 937209"
/tissue_type="muscle"
/dev_stage="adult"
/lab_host="SOLR (Kanamycin resistant)"
/note="Organ: skeletal muscle; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally.
Primer: Oligo dt. Skeletal muscle from patient with malignant hyperthermia. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTT 3' "

BASE COUNT 97 a 78 c 77 g 124 t
ORIGIN

Query Match 74.8%; Score 17.2; DB 30; Length 376;
Best Local Similarity 86.4%; Pred. No. 66;


```

ACCESSION      IMAGE:635002.5', mRNA sequence.
NID            AA164056
VERSION        91740009
KEYWORDS       AA164056.1 GI:1740009
SOURCE         EST.
ORGANISM       house mouse.
AUTHORS        Mus musculus.
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1 (bases 1 to 465)
AUTHORS        Marra,M., Hillier,L., Allen,M., Bowles,M., Dierlich,N., Dubuque,T.,
               Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Thelning,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE          The WashU-HHMI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        On Jan 25, 1995 this sequence version replaced gi:637825.

FEATURES             location/qualifiers
     source           1..465
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /map="22"
                     /clone="IMAGE:635002"
                     /clone_1lb="Soares mouse lymph node NBMUN"
                     /sex="male"
                     /tissue_type="lymph node"
                     /dev_stage="4 weeks"
                     /lab_host="DH10B"
                     /note="Organ: lymph node; Vector: pTZ19D-Pac (Pharmacia)
                     with a modified polylinker; Site.1: Not I; Site.2: Eco RI;
                     1st strand cDNA was primed with a Not I - oligo(dT) primer;
                     15' strand cDNA was primed with a Not I - oligo(dT) primer."
TGTTACCATGTGAAGTGGGAGCGCCGCATCTTTTTTTTTTTTTTTTTTTTT
3'] : double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pTZ19 vector. RNA
provided by Dr. Bertrand Jordan, library constructed and
normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      140 a          94 c          82 g          149 t
ORIGIN
Query Match    73.0%; Score 16.8; DB 29; Length 465;
Best Local Similarity   90.0%; Pred. No.1.1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY              3 TTCACACTGGCATGCCGA 22
                ||||| ||||||||| |
Db              329 TTCACAGACTGGCATGCCAA 310

RESULT 11
LOCUS          AA272905              376 bp      mRNA
DEFINITION     vas9d01.r1 Scores mouse J3NEI2.5 Mus musculus cDNA clone
ACCESSION      AA272905
NID            91907497
VERSION        AA272905.1 GI:1907497
```

```

KEYWORDS      EST.
SOURCE         house mouse.
ORGANISM       Mus musculus
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               1 (passes 1 to 376)
               Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
               Geisel,S., Kuchta,T., Lacy,M., Le,M., Martin,J., Morris,M.,
               Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
               Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
               Waterston,R.
TITLE          The WashU-HMNI Mouse EST Project
JOURNAL        Unpublished (1996)
COMMENT        On Sep 12, 1996 this sequence version replaced gi:1394067.

Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watsn.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MEI:450777
Seq primer: -28mj3 rev2 ET from Amersham
High quality sequence stop: 361.
Location/Qualifiers
1. 376
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:733729"
/clone_1lb="Soares mouse 3NNE12 5"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH10B"
/note="Organ: whole fetus; Vector: pT773D-Pec (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTCCACATCTGAAGTCGACGCGCCGCCCTATTTTCTTTTCTTTT
3'], on total mouse RNA [provided by Minoru Ko, Wayne
State Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT773 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT    128 a 77 c 78 g 93 t
ORIGIN
Query Match      73.0%; Score 16.8; DB 30; Length 376;
Best Local Similarity 90.0%; Pred. No. 1e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0
QY             2 ATTGCACACTGCATGCGCG 21
Db            81 ATTGCACACTGCATGCTG 100

RESULT 12
LOCUS      AA273650          398 bp      mRNA      EST      28-MAR-1997
DEFINITION vD92c09.r1 Soares mouse 3NBMS Mus musculus cDNA clone IMAGE:764464
ACCESSION  AA273650
NID        91912717
VERSION    AA273650.1  GI:1912717
KEYWORDS   EST
SOURCE     house mouse.
ORGANISM   Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```


FEATURES
source
1..638
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_type="oocyte"
107..583
/gene="ubc9"
107..583
/gene="ubc9"
/note="ubc9"
/codon_start=1
/product="E2 ubiquitin conjugating enzyme"
/protein_id="AA57736.1"
/db_xref="GI:2078331"
/translation="MSGIALSLAOKERKARKDPGFVAVPTKNDGTMINLMECA
IPGKKGWGGGLFKLMLEKDDYPPSPKCKPEPLFHNVPSTGVCISIEEDK
MRPAITIKQLLQIOLNEPNIODPAQAEAYITTCQNRVEYKRVRAQAKKFAFS"
BASE COUNT 214 a 131 c 147 g 146 t
ORIGIN

Query Match 100.0%; Score 35; DB 4; Length 638;
Best Local Similarity 100.0%; Pred. No. 0.078;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGTGT 35
|||||
DB 638 GACTCGAGTCGACATCGATTTTGTGTGT 604

RESULT 2
A24382 35 bp DNA PAT 21-MAR-1995
LOCUS d517 adaptor primer.
DEFINITION A24382
ACCESSION A24382
NID 9904477
VERSION A24382.1 GI:904477
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 35)
AUTHORS Sermentos, P.; De Taxis du Poet, P.; Nittl, G. and Scacheri, E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: EP 0501821 A 22 02-SEP-1992;
FARMITALIA CARLO ERBA S.R.L.
FEATURES
source 1..35
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGT 35
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTGT 35

RESULT 3
A27645 35 bp DNA PAT 04-JUN-1995
LOCUS Hybrid d517-adaptor primer.
DEFINITION A27645
ACCESSION A27645
NID 91248481
VERSION A27645.1 GI:1248481
KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 35)
AUTHORS
TITLE POLYPHENOL OXIDASE GENES
JOURNAL Patent: WO 9302195-A 1 04-FEB-1993;
FEATURES
source 1..35
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGT 35
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTGT 35

RESULT 4
A37244 35 bp DNA PAT 05-MAR-1997
LOCUS Sequence 13 from Patent WO9403609.
DEFINITION A37244
ACCESSION A37244
NID 92294355
VERSION A37244.1 GI:2294355
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Parker, P.J.; Goode, N.T.; Nurse, P.M. and Waterfield, M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 13 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTGTGTGT 35
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTGT 35

RESULT 5
A40373 35 bp DNA PAT 05-MAR-1997
LOCUS Sequence 29 from Patent WO9425489.
DEFINITION A40373
ACCESSION A40373
NID 92296422
VERSION A40373.1 GI:2296422
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Mohapatra, S.S. and Sehon, A.H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL Patent: WO 9425489-A 29 10-NOV-1994;

UNIV MANITOBA (CA)
 COMMENT Other publication AU 6674094 941121.
 FEATURES
 source 1. .35
 /db_xref="taxon:32644"
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 6
 A42335 35 bp DNA PAT 05-MAR-1997
 LOCUS
 DEFINITION Sequence 7 from Patent WO9502057.
 ACCESSION A42335
 NID 92297812
 VERSION A42335.1 GI:2297812
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
 Kamalati,T., Page,M.J. and Spence,P.
 TITLE PROTEIN TYROSINE KINASE AND LIGANDS THEREOF
 JOURNAL Patent: WO 9502057-A 7 19-JAN-1995;
 COMMENT Other publication AU 7080994 950206.
 FEATURES
 source 1. .35
 /db_xref="taxon:32644"
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 7
 A42384 35 bp DNA PAT 06-MAR-1997
 LOCUS
 DEFINITION Sequence 7 from Patent WO9502187.
 ACCESSION A42384
 NID 92297858
 VERSION A42384.1 GI:2297858
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
 Martindale,J.E., Page,M.J. and Spence,P.
 TITLE CELL GROWTH FACTOR RECEPTORS
 JOURNAL Patent: WO 9502187-A 7 19-JAN-1995;
 COMMENT Other publication AU 7081094 950206.
 FEATURES
 source 1 (bases 1 to 35)
 /db_xref="taxon:32644"

source 1. .35
 /db_xref="taxon:32644"
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 8
 A46467 35 bp DNA PAT 07-MAR-1997
 LOCUS
 DEFINITION Sequence 5 from Patent WO9526402.
 ACCESSION A46467
 NID 92300644
 VERSION A46467.1 GI:2300644
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Knox,D.P., Smith,S.K., Smith,W.D., Redmond,D. and Murray,J.
 TITLE VACCINES AGAINST HELMINTHIC PARASITES
 JOURNAL Patent: WO 9526402-A 5 05-OCT-1995;
 COMMENT MALINKRODT VETERINARY INC (US)
 Other publication CA 2182178 951005
 Other publication AU 1956495 951017.
 FEATURES
 source 1. .35
 /db_xref="taxon:32644"
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 9
 A59198 40 bp DNA PAT 06-MAR-1998
 LOCUS
 DEFINITION Sequence 7 from Patent WO9704108.
 ACCESSION A59198
 NID 93714579
 VERSION A59198.1 GI:3714579
 KEYWORDS
 SOURCE unidentified.
 ORGANISM unidentified.
 REFERENCE 1 (bases 1 to 40)
 AUTHORS Schuster,E., Spiessler,B., Titzel,K., Gottschalk,M., Khan,N.Q.,
 Wolf,S. and Plainer,H.
 TITLE LEUCINE AMINOPEPTIDASES PRODUCED RECOMBINANTLY FROM ASPERGILLUS
 JOURNAL Patent: WO 9704108-A 7 06-FEB-1997;
 COMMENT ROEHM GMBH (DE)
 Other publication DE 19526485 970123.
 FEATURES
 source 1. .40
 /db_xref="taxon:32644"

LOCUS 128284 35 bp DNA PAT 30-OCT-1996
 DEFINITION Sequence 13 from patent US 5569830.
 ACCESSION 128284
 MID g1819060
 VERSION 128284.1 GI:1819060
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE Unclassified.
 1 (bases 1 to 35)
 AUTHORS Bennett,A., Labavitch,J.M., Powell,A. and Stoltz,H.
 TITLE Plant inhibitors of fungal polygalacturonases and their use to
 control fungal disease
 JOURNAL Patent: US 5569830-A 13 29-OCT-1996;
 FEATURES Location/Qualifiers
 source 1..35
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35
 Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

RESULT 15
 134242
 LOCUS 134242 35 bp DNA PAT 30-JAN-1997
 DEFINITION Sequence 19 from patent US 5597569.
 ACCESSION 134242
 MID g1825033
 VERSION 134242.1 GI:1825033
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE Unclassified.
 1 (bases 1 to 35)
 AUTHORS Siegfall,C.B., Gawlak,S.L. and Marguardt,H.
 TITLE Bryodin 2 a ribosome-inactivating protein isolated from the plant
 Bryonia dioica
 JOURNAL Patent: US 5597569-A 19 28-JAN-1997;
 FEATURES Location/Qualifiers
 source 1..35
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 100.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.095;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35
 Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

Search completed: September 13, 1999, 15:55:19
 Job time: 4547 sec

THIS PAGE BLANK (USPTO)

KM polymerase chain reaction; amplify; Haemonchus contortus; ss.
 OS Synthetic.
 PN WO9323542-A.
 PD 25-NOV-1993.
 PE 07-MAY-1993: G00943.
 PR 08-MAY-1992: GB-0009993.
 (AGRI-) AGRIC & FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;
 PI Smith TS;
 DR WPI: 93-386574/48.
 PT New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 PS Example; Page 32: 137pp; English.
 CC The sequences given in 052503-11 are primers which were used to
 CC amplify cDNA clones of the helminth aminopeptidase genes H11-1, -2
 CC and -3. The amplified sequences encode H110D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between the amplified clones can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H110D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in
 CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by the H11 genes may be used in vaccines to stimulate immune
 CC response against helminth parasites in humans or other animals. These
 CC DNA sequences may be incorporated into a virus or microbe and used in
 CC a similar manner.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;
 Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GACTCGAGTCGACATCGATTTT 35
 1 GACTCGAGTCGACATCGATTTT 35
 RESULT 3
 ID 059015
 AC 059015: standard; DNA: 35 BP.
 DT 12-APR-1994 (first entry)
 DE RACE PCR adaptor-dt.
 KW phosphonositide kinase; PI: p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphonositides;
 KW ss; amplification; p110.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PE 13-APR-1993: G00761.
 PR 13-APR-1992: GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parnerp, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphonositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 39: 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dt as primers. Products were fractionated using an agarose gel. The
 CC DNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also 051155-6, 059012-23 and 057522-3.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GACTCGAGTCGACATCGATTTT 35
 1 GACTCGAGTCGACATCGATTTT 35
 RESULT 4
 ID 057021
 AC 057021: standard; DNA: 35 BP.
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase adaptor dt primer.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schistosomacromyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PE 05-AUG-1993: G01651.
 PR 05-AUG-1992: GB-016554.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PUJ, Waterfield MD;
 DR WPI: 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 15: 71pp; English.
 CC The sequences given in 057020-23 are primers which were used in the
 CC isolation of the phosphatidylinositol (Ptdins) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schistosomacromyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the Ptdins is induced. Ptdins activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;
 Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 GACTCGAGTCGACATCGATTTT 35
 1 GACTCGAGTCGACATCGATTTT 35
 RESULT 5
 ID 071021
 AC 071021: standard; DNA: 35 BP.
 DT 26-SEP-1994 (first entry)
 DE Primer for amplifying peach polygalacturonase coding sequence.
 KW Ripening; melting; pectins; breeding; Prunus; Rosaceae;
 KW endopolygalacturonase; EndoPG; polygalacturonase; PG; ss.
 OS Synthetic.
 PN WO9405795-A.
 PD 17-MAR-1994.
 PE 27-AUG-1993: AU0442.
 PR 27-AUG-1992: AU-004370.
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (PACI-) PACIFIC SEEDS PRY LTD.
 PI Brady CJ, Lee E, Lester DR, Orr G, Speirs J;

DR WPI: 94-101199/12.
 PT DNA encoding peach polygalacturonase (PG) and production of
 PT recombinant peach PG in rosaceae plants - for production of
 PT plants with 'melting' or 'non-melting' phenotype
 PS Disclosure; Page 15; 51pp; English.
 CC During early ripening of peaches, tissue firmness decreases
 CC slowly and progressively. Towards the end of ripening, loss of
 CC tissue firmness is rapid. This second stage of softening is called
 CC the "melting" stage. Fruit of peach varieties used for canning do
 CC not have a "melting" phase of softening. Ripe fruit remain
 CC relatively firm and maintain their shape throughout processing.
 CC Fruit of "melting" varieties show an increase in activity of
 CC endopolygalacturonase (EndoPG) during ripening. Polygalacturonase
 CC (PG) is believed to contribute to fruit softening through its action
 CC on intercellular and cell wall pectins. This primer was used to
 CC synthesize a first strand cDNA molecule for further amplification
 CC by PCR. See 056765-67.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 DB 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 6
 063873
 ID 063873 standard; cDNA: 35 BP.
 AC 063873:
 TM 04-NOV-1994 (first entry)
 DE T cell protein CD4 adaptor-dt17 primer.
 KW Cat; feline immunodeficiency virus; CD4 positive T lymphocyte;
 KW cellular receptor; T cell protein; polymerase chain reaction; PCR; ss.
 OS Synthetic.
 PN FR2696188-A.
 PD 01-APR-1994.
 TM 25-SEP-1992; 011496.
 PR 25-SEP-1992; FR-011496.
 PA (UYPA-) UNIV CURIE PARIS VI P 6 M.
 PI De Parseval A, Klatzmann D, Salmon P, Sonigo P;
 DR WPI: 94-128285/16.
 PI Feline CD4 protein and corresp. DNA - for study of feline
 PI immunodeficiency virus and for therapy
 PS Disclosure; Page 5; 25pp; French.
 CC The primers 063871 and 063872 were used to obtain cDNA coding for
 CC feline CD4 protein by polymerase chain reaction. For rapid
 CC amplification of the ends of the cDNA, a hybrid adaptor-oligo(dt)
 CC primer was also used. Cat lymph node cDNA was amplified and the
 CC full-length CD4 coding sequence 063870 was obtained from several
 CC partial clones.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 DB 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 7
 065594
 ID 065594 standard; DNA: 35 BP.
 AC 065594:
 DT 12-JAN-1995 (first entry)
 DE HER4 primer XSC17.
 KW Human epidermal growth factor receptor; HER; EGF; tyrosine kinase;

KW cancer; neuronal tissue; muscle tissue; neoplasm;
 KW carcinoma; primer; probe; PCR; ss.
 OS Synthetic.
 PN EP-599274-A.
 PD 01-JUN-1994.
 PF 23-NOV-1993; 118837.
 PR 24-NOV-1992; US-981165.
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI Culoscou J, Plowman GD, Shoyab M;
 DR WPI: 94-169599/21.
 PT New recombinant nucleic acid expressing HER4 - a new receptor
 PT tyrosine kinase expressed in some cancer cells, and related
 PT vectors, antibodies, ligands etc, for diagnosis and treatment of
 PT cancers.
 PS Example; Page 16; 104pp; English.
 CC The primers given in 065594-6 were used in the semi-quantitative
 CC PCR detection of HER4.
 CC HER4 is the fourth member of the EGFR-family of tyrosine kinases and
 CC is expressed in some human cancers and in some tissues of neuronal
 CC or muscle origin. HER4 polynucleotides, opt. labelled, are useful
 CC in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
 CC breast carcinoma) and as primers in PCR or as probes.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 DB 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 8
 078807
 ID 078807 standard; cDNA: 35 BP.
 AC 078807:
 DT 12-JUN-1995 (first entry)
 DE TCR alpha-chain dt17 adaptor.
 KW Allergen; immunotherapy; T-cell receptor; TCR; CDR3;
 KW complementarity determining region-3; immunosuppressive;
 KW antibody engineering; suppressor T-lymphocyte; Ts; PCR;
 KW polymerase chain reaction; amplification; adaptor; ss.
 OS Synthetic.
 PN WO9425489-A.
 PD 10-NOV-1994.
 PF 26-APR-1994; CA0228.
 PR 26-APR-1993; GB-008381.
 PA (UYMA-) UNIV MANITOBA.
 PI Mohapatra SS, Sehon AH;
 DR WPI: 94-358193/44.
 PT Synthetic peptide(s) for effecting immuno-therapy - have an amino
 PT acid sequence comprising a portion of the CDR3 region of a T-cell
 PT receptor.
 PS Disclosure; Page 43; 78pp; English.
 CC Antibody production against allergenic antigen (Ag) is specifically
 CC suppressed by treatment with the junctional segment of the alpha
 CC and/or beta chain of the Ag receptor of suppressor T (Ts) cells
 CC (TCR) induced by tolerogenic Ag-PBS conjugates. cDNAs encoding
 CC the alpha and beta chains of TCRs of cloned Ts cells specific for
 CC ovalbumin and human monoclonal (myeloma) IgG were produced by PCR
 CC using the primers given in 078798-805 and the adaptors given in
 CC 078806-07.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.011;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35
 DB 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 9

ID 084786 standard; DNA; 35 BP.

AC 084786

DT 17-AUG-1995 (first entry)

DE PTK22 reverse-transcription primer.

KW protein-tyrosine-kinase; PTK; discoidin domain receptor; cancer;

KM breast tumor; mamma carcinoma; diagnosis; prognosis; therapy; ss.

OS Synthetic.

PN WO9502187-A.

PD 19-JAN-1995.

PF 08-JUL-1994; G01480.

PR 09-JUL-1993; GB-014271.

PA (CANC-) CANCER RES INST.

PI (WELL) WELLCOME FOUND LTD.

PI Barker KT, Crompton MR, Gusterson BA, Martindale JE;

PI Mitchell PJ, Page MJ, Spence P;

DR WPI: 95-066991/09.

PT Method for screening substances, using protein tyrosine kinase -

PT for potential utility as therapeutic agents for cancer

PS Disclosure; Page 34; 51pp; English.

CC CDNA derived from tumor metastatic tissue was amplified using

CC primers (given in 084783-84) based on sequences (R71101, R71103)

CC associated with protein-tyrosine-kinases (PTK). Novel PTK22 was

CC identified in an isolated subclone. The 3' sequence of PTK22 was

CC obtained by reverse transcription (using the primer of 084786) and

CC PCR amplification (primers 084787-88) of RNA of human breast

CC carcinoma cell line MDA MG 468. The partial DNA sequence of PTK22

CC is given in 084782.

SC Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Query Match 100.0%; Score 35; DB 1; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC 090167 standard; DNA; 35 BP.

AC 090167;

DT 01-NOV-1995 (first entry)

DE Pollen allergen gene primer 8.

KW Japanese cedar; pollen; allergen; allergy; therapy; diagnostic;

KM desensitizer; Cryptomeria japonica; polymerase chain reaction;

KW PCR; primer; ss.

OS Synthetic.

PN EP-655500-A.

PD 31-MAY-1995.

PF 03-NOV-1994; 308117.

PR 05-NOV-1993; JP-289151.

PR 20-DEC-1993; JP-344596.

PR 27-DEC-1993; JP-346814.

PA (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.

PI Kurimoto M, Namba M, Torisoe K;

PI WPI: 95-195588/26.

DR New Japanese cedar pollen allergen polypeptide - and DNA coding for it,

PT useful for treatment and diagnosis of cedar pollen allergy

PS Disclosure; Page 11; 41pp; English.

CC The primers given in 090166-69, based on a portion of cDNA clone

CC SC09 or on previously isolated sequences, were used for the PCR

CC amplification of cDNA derived from Japanese cedar pollen, to obtain

CC clone SC50 bearing a partial sequence (nt 1-198 of the sequence

CC given in 090156) of the pollen allergen gene.

SC Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 11

ID 089878 standard; DNA; 35 BP.

AC 089878;

DT 31-JAN-1997 (Revised)

DT 25-OCT-1995 (first entry)

DE PCR primer for cloning bryodin-2.

KW Ribosome inactivating protein; cell death; proliferation; tumour;

KM cancer; virus; malaria; trypanosomiasis; inflammation; autoimmune;

KW disease; purging; bone marrow; ss.

OS Synthetic.

PN WO9511977-A.

PD 04-MAY-1995.

PF 25-OCT-1994; U12382.

PR 25-OCT-1993; US-141891.

PR 20-OCT-1994; US-324301.

PA (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Gawlak SL, Marquardt H, Siegal CB;

DR WPI: 95-178870/23.

PT New ribosome inactivating protein bryodin-2 and related

PT conjugates - for killing target cells, e.g. in treatment of

PT cancer, also related nucleic acid, vectors and transformed cells

PS Disclosure; Page 43; 81pp; English.

CC The sequence is that of a PCR primer used to isolate a ribosome

CC inactivating protein termed bryodin-2, from Bryonia dioica. Conjugates

CC of the protein with a ligand may be used to kill target cells (i.e. those

CC to which the ligand component binds specifically) or to inhibit

CC proliferation of tumour cells. Typical applications include the

CC treatment of cancer, viral infections, malaria, trypanosomiasis,

CC inflammatory or autoimmune diseases, including in vitro purging of e.g.

CC bone marrow.

CC See also 089874-86.

CC Updated 31-JAN-1997 due to incorrect cross reference numbers.

SC Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Query Match 100.0%; Score 35; DB 1; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.011;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 12

T05321 standard; DNA; 35 BP.

AC T05321;

DT 13-APR-1996 (first entry)

DE Oligonucleotide probe for human prostacyclin-synthase.

KW DNA probe; prostacyclin-synthase; PCR; polymerase chain reaction;

KM DNA primer; prostaglandin I2; circulatory disease; therapeutic;

KW diagnosis; gene therapy; ss.

OS Synthetic.

PN WO9530013-A1.

PD 09-NOV-1995.

PF 27-APR-1995; J00838.

PR 28-APR-1994; JP-114316.

PA (TANA/) TANABE T.

PI Tanabe T;

DR WPI: 95-393084/50.

PT Human prostacyclin synthase and DNA encoding it - useful in the

Investigation and treatment of diseases characterised by reduced prostaglandin 12 production.

PS Disclosure; Page 34; 71pp; Japanese.

CC DNA primers (T05317-20; T05322; T05326-27) are used to screen human genomic lung cell line W138 and human arterial endothelial cell CC cDNA libraries for the isolation of a prostacyclin-synthase (PGIS) CC coding sequence (see T05316). DNA probes (T05321 and T05323) are used CC in the construction of plasmid pHPG1S1, encoding the complete PGIS CC sequence. This plasmid was used to transfect human 293 cells for CC PGIS peptide expression. DNA encoding human PGIS, vectors containing CC 1 α , and PGIS itself, may be administered to patients to increase CC prostaglandin 12 (PGI₂) production to treat diseases characterized CC by reduced PGI₂ levels or by an imbalance between PGI₂ and CC thromboxane A₂ levels, such as circulatory diseases (thrombosis, CC angina pectoris, arteriosclerosis, myocardial infarction). The CC cDNA and protein are also useful in disease diagnosis.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 13

ID 094244 standard; DNA: 35 BP.

AC 094244;

DT 14-MAY-1996 (first entry)

DE Cysteine proteinase active chain antisense polyA primer.

KW Primer; amplify; polymerase chain reaction; PCR; Haemonchus contortus;
KW cysteine proteinase; DM.2; DM.3; DM.4; human; DM.4a; DM.5; antigen;
KW vaccine; helminth; parasite; DM.1; ruminant; gastro-intestinal tract; ss.
OS Synthetic.

PN W09526402-A1.

PD 05-OCT-1995.

PF 24-MAR-1995; G00665.

PR 25-MAR-1994; GB-005990.

PR 25-MAR-1994; GB-005925.

PA (MCM) MALLINCKRODT VETERINARY INC.

PI Knox DP, Murray J, Redmond D, Smith SK, Smith WD;

DR WPI; 96-351322/45.

PT Protective helminth parasite antigen - used in vaccine directed PT against parasitic nematodes of mammalian gastro-intestinal tract

PS Example 16; Fig 15; 79pp; English.

CC The sequences given in 094240-45 are primers which are based on CC previously published sequences derived from the canonical Haemonchus CC contortus cysteine proteinase molecule. These primers were used in CC the cloning of cDNA fragments from the cysteine proteinase gene, CC such as DM.1, DM.2, DM.3, DM.4, DM.4a and DM.5 (see also 094246-51).

CC The amplified fragments may be expressed in a recombinant cell for CC the production of antigens. These antigens may be used in the CC preparation of a vaccine against helminth parasites in a human or CC non-human animal.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 14

T10696

Investigation and treatment of diseases characterised by reduced prostaglandin 12 production.

PS Disclosure; Page 34; 71pp; Japanese.

CC DNA primers (T05317-20; T05322; T05326-27) are used to screen human genomic lung cell line W138 and human arterial endothelial cell CC cDNA libraries for the isolation of a prostacyclin-synthase (PGIS) CC coding sequence (see T05316). DNA probes (T05321 and T05323) are used CC in the construction of plasmid pHPG1S1, encoding the complete PGIS CC sequence. This plasmid was used to transfect human 293 cells for CC PGIS peptide expression. DNA encoding human PGIS, vectors containing CC 1 α , and PGIS itself, may be administered to patients to increase CC prostaglandin 12 (PGI₂) production to treat diseases characterized CC by reduced PGI₂ levels or by an imbalance between PGI₂ and CC thromboxane A₂ levels, such as circulatory diseases (thrombosis, CC angina pectoris, arteriosclerosis, myocardial infarction). The CC cDNA and protein are also useful in disease diagnosis.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.011;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35
DB 1 GACTCGAGTCGACATCGATTTTATTTTATTTT 35

RESULT 15

ID T10696 standard; DNA: 35 BP.

AC T10696;

DT 26-MAY-1996 (first entry)

DE Trypsin-like enzyme oligonucleotide TRX-10.

KW Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW protease; primer; PCR; amplification; ds.
OS Synthetic.

PN AU9527248-A.

PD 08-FEB-1996.

PF 31-JUL-1995; 027248.

PR 29-JUL-1994; JP-178607.

PA (TEIJ) TEIJIN LTD.

PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;

PI Yamaoka K, Yamaoka S;

DR WPI; 96-117356/13.

PT Nucleic acid sequence encoding trypsin-like enzyme - which digests PT fibrinogen, used as expectorant in treatment of respiratory PT diseases, e.g. bronchial asthma

PS Example 9; Page 44; 65pp; English.

CC Example 9 describes the cloning of cDNA region encoding CC trypsin-like enzyme isolated from cough phlegm.

CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8 CC (T10695), TRX-10 (T10696) and TRX-11 (T10697).

CC TRX-1 corresponds to from A1 to A23 of the sequence given in CC T10693, which is part of a cDNA encoding the trypsin-like CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OK nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:55 ; Search time 1694.61 Seconds
(without alignments)
40.740 Million cell updates/sec

Title: US-09-325-095-16

Perfect score: 35
Sequence: 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTTTTTT 35

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	35	100.0	337	32	AA340759	AA340759 EST46034
2	25.6	73.1	678	44	A1295887	A1295887 LP09649.5
3	25.4	72.6	111	44	A1312326	A1312326 ta7Be03.x
4	25.4	72.6	380	45	A1348883	A1348883 tb05c11.x
5	25.4	72.6	909	48	A1568751	A1568751 th15d09.x
6	25.4	72.6	466	51	A1717729	A1717729 UI-R-YO-a
7	24.8	70.9	305	23	L38011	L38011 BNAF0228E M
8	24.8	70.9	435	28	C18175	C18175 C18175 Huma
9	24.6	70.3	314	35	AA541358	AA541358 n177d04.s
10	24.6	70.3	466	51	A1726205	A1726205 BNLGH1525
11	24.6	70.3	623	51	A1729852	A1729852 BNLGH1535
12	24.4	69.7	405	35	AA541329	AA541329 n176e11.s
13	24.4	69.7	511	51	A1732239	A1732239 nf35f10.x
14	24.4	69.7	517	51	A1734188	A1734188 nc78e12.y
15	24.2	69.1	570	29	AA145612	AA145612 ms07h10.T
16	24.2	69.1	214	38	AA751165	AA751165 ISM40847
17	24.2	69.1	656	44	A1322793	A1322793 ms07h10.Y
18	24.2	69.1	468	47	A1517504	A1517504 GH28493.5
19	24	68.6	168	43	A1223480	A1223480 qx27b01.x
20	24	68.6	149	43	A1223537	A1223537 qx28h03.x
21	24	68.6	146	43	A1224352	A1224352 qx99d09.x
22	24	68.6	64	43	A1224631	A1224631 qx97e03.x
23	24	68.6	92	44	A1246064	A1246064 qx52e10.x
24	24	68.6	143	44	A1250505	A1250505 qx33a05.x
25	24	68.6	137	44	A1250569	A1250569 qx31g07.x
26	24	68.6	267	44	A1251260	A1251260 qx36g05.x
27	24	68.6	215	44	A1251658	A1251658 qx32b04.x
28	24	68.6	90	44	A1252577	A1252577 qx25a08.x
29	24	68.6	239	44	A1252739	A1252739 qx26b09.x
30	24	68.6	360	44	A1254267	A1254267 qx33f07.x
31	24	68.6	190	44	A1254390	A1254390 qx33d05.x
32	24	68.6	177	44	A1254415	A1254415 qx34b11.x
33	24	68.6	262	44	A1254455	A1254455 qx47f11.x
34	24	68.6	175	44	A1254797	A1254797 qx53d08.x
35	24	68.6	396	44	A1254907	A1254907 qx55a02.x
36	24	68.6	190	44	A1255106	A1255106 qx46a04.x
37	24	68.6	176	44	A1266290	A1266290 qx69a05.x
38	24	68.6	146	44	A1271282	A1271282 qx83h03.x
39	24	68.6	160	44	A1271407	A1271407 qx82e11.x
40	24	68.6	257	44	A1284677	A1284677 qx24a08.x
41	24	68.6	61	44	A1284843	A1284843 qx25d10.x
42	24	68.6	61	44	A1305311	A1305311 qx76f07.x
43	24	68.6	156	44	A1305926	A1305926 qx79a04.x
44	24	68.6	126	44	A1306336	A1306336 qx80b09.x
45	24	68.6	109	44	A1306853	A1306853 qx70c06.x

ALIGNMENTS

RESULT 1
AA340759
LOCUS
DEFINITION
ACCESSION
NID
VERSION

AA340759 337 bp mRNA
EST46034 Fetal kidney II Homo sapiens cDNA 5' end similar to
similar to protein phosphatase 2A, beta, 55 kDa, mRNA sequence.
AA340759
g1992998
AA340759.1 GI:1992998

	KEYWORDS	EST.
SOURCE	ORGANISM	human.
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	REFERENCE	Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS		Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fullmer,R.A.,
		Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
		White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man'hai,C,
		Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
		Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghegan,N.S.,
		Glotzer,A.O., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
		Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
		Morono-Palankues,R.F., McDonald,J.A., Nguyen,D.T., Pelligrino,S.M.,
		Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R.,
		Small,K.V., Springs,T.A., Utterback,T.R., Weidman,J.F., Wix, *Bedarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.O., Dimke,D., Feng,D.-F., Ferlie,J.A., Fischer,C., Hastings,G.A., He'M.W., Hu,Y.D., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., HungJun,J., Li,H., Meisner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wang,J., Xu,C., Yu,G.L., Rubin,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence Nature 377 (6547 Suppl.), 3-174 (1995) On Sep 12, 1996 this sequence version replaced gi:1406936. Other ESTs: EST46035 THC125647 Contact: Kerlavage, AR Bioinformatics for Genomic Research The Institute for Genome Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel.: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the tIGR Human Gene Index (http://www.tigr.org/tcdb/hgi/hgi.html) Seq primer: M13-21.
FEATURES	Source	Location/Qualifiers
		1..337
		/organism="Homo sapiens"
		/db_xref="ATCC (Inhost):142407"
		/db_xref="taxon:9606"
		/clone_lib="fetal kidney II"
		/dev_stage="fetufus"
		/note="Organ: kidney; Vector: pBluescript KS-; Site_1: XhoI; Site_2: EcoRI"
	BASE COUNT	63 a 88 c 95 g 88 t 3 others
	ORIGIN	
	Query Match	100.0%; Score 35; DB 32; Length 337; Best Local Similarity 100.0%; Pred. No. 1.1; Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1	GACTCGAGTGCACATCGATTCTTTTCTTTTTTTT 35 A1295887
Db	4	GACTCGAGTCGACATCGATTCTTTTCTTTTTTTT 38
RESULT	2	
LOCUS	A1295887/c	678 bp mRNA EST P1-DEC-1998
LINITION	LP09649.5prline LP Drosophila melanogaster larval-early pupal port2	
ACCSSION	Drosophila melanogaster CDNA Clone LP09649 5prime, mRNA sequence.	
NID	A1295887	
VERSION	g3945294	
KEYWORDS	A1295887.1 GI:3945294	
SOURCE	EST.	tissue fly.
ORGANISM	Drosophila melanogaster	

[illegible]

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/dbcrp/image/Image.html

Seq primer: -40UP from Gldco.

FEATURES
SOURCE
1. .111
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2050204"
/clone_lib="NCI CGAP_HSC2"
/tissue_type="stem cell 34+/38+"
/dev_stage="adult"
/lab_host="DH10B"

/note="Organ: Bone marrow; Vector: PAMPI; mRNA made from bone marrow, stem cells 34+/38+, cDNA made by oligo-dt priming. Directionally cloned. Size selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

BASE COUNT
ORIGIN
71 a 10 c 22 g 8 t

Query Match
Best Local Similarity 72.6%; Score 25.4; DB 44; Length 111;
Matches 29; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35
||||| ||||| ||||| ||||| ||||| |||||
Db 66 GACTCGATACGACTTTTCTTTTCTTTTCTTTT 32

LOCUS
AI348883 380 bp mRNA EST 16-FEB-1999
DEFINITION
tld05c11.x2 NCI CGAP lu26 Homo sapiens cDNA clone IMAGE:2052692 3'
similar to contans Alu repetitive element; contains LI.12 LI
repetitive element ;, mRNA sequence.

ACCESSION
NID
94086089
AI348883
AI348883.1 GI:4086089

KEYWORDS
SOURCE
human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
COMMENT
On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

cDNA Library Preparation: David B. Krizman, Ph.D.

cDNA Library Arrayed by: I.M.A.G.E. Consortium, LLNL

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbcrp/image/Image.html

Insert Length: 452 Std Error: 0.00
Seq primer: -40UP from Gldco.

FEATURES
SOURCE
Location/Qualifiers

1. .380

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:2052692"

/clone_lib="NCI CGAP lu26"

/tissue_type="invasive adenocarcinoma"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: Lung; Vector: PAMPI; mRNA made from lung adenocarcinoma tissue, cDNA made by oligo-dt priming. Directionally cloned. Size selected on agarose gel, average insert size 500 bp. Primary library, non-amplified."

BASE COUNT
ORIGIN
106 a 68 c 71 g 135 t

Query Match
Best Local Similarity 82.9%; Score 25.4; DB 45; Length 380;
Matches 29; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35
||||| ||||| ||||| ||||| ||||| |||||
Db 26 GACTCGATACGACTTTTCTTTTCTTTTCTTTT 60

RESULT 5
AI568751 909 bp mRNA EST 29-MAR-1999
LOCUS
DEFINITION
th15d09.x1 NCI CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2118353 3',
mRNA sequence.

ACCESSION

NID

AI568751 94532125
AI568751.1 GI:4532125

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

TITLE

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL

COMMENT

On Feb 18, 1999 this sequence version replaced gi:4297567.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,

M.D., Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbcrp/image/Image.html

Seq primer: -40UP from Gldco

High quality sequence stop: 448.

Location/Qualifiers

1. .909

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="11p11"

/clone="IMAGE:2118353"

/clone_lib="NCI CGAP_CLL1"

/tissue_type="B-cell, Chronic lymphocytic leukemia"

/lab_host="DH10B"

/note="Vector: pRT3D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAGTGGAGCGCGCATGCTTTTCTTTTCTTTTCTTTT

T 3']; double-stranded cDNA was ligated to Eco RI

adaptors (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of the modified pRT3D vector.

Library is normalized, and was constructed by Bento

Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
186 a 224 c 250 g 244 t 5 others

ACCESSION C18175
 VERSION g1579777
 C18175.1 GI:1579777
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 435)
 Fujitawa, T., Hirano, H., Hishigaki, H., Horie, M., Kawai, A., Kuga, Y., Kyushiki, H., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomura, H., Suzuki, M., Takai, A., Takeda, S., Matanabe, T., Makawa, H., Nakamura, Y. and Takahashi, E.
 TITLE Otsuka cDNA project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1392743.
 CONTACT: Tsutomu Fujitawa
 Otsuka GEN Research Institute
 Otsuka Pharmaceutical Co., Ltd
 463-10 Kagasuno Kawasuchi-cho, Tokushima, Tokushima, 771-01 Japan
 Tel: 0886-65-2888
 Fax: 0886-37-1035.
 FEATURES Location/Qualifiers
 source 1..435
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="GEN-558E05"
 /clone_lib="Human placenta cDNA (Fujitawa)"
 /tissue_type="placenta"
 BASE COUNT 131 a 88 c 111 g 105 t
 ORIGIN
 Query Match 70.9%; Score 24.8; DB 28; Length 435;
 Best Local Similarity 92.9%; Pred. No. 3.8e+02;
 Matches 26; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Oy 8 GTCGACATCGATTTTCTTTT 35
 ||||| ||||| ||||| ||||| |||||
 Db 274 GTCGACCTCGAGTTTCTTTT 247
 RESULT 9
 AA541358 314 bp mRNA EST 19-AUG-1997
 LOCUS n177d04.s1 NCI_CGAP_Pri2 Homo sapiens cDNA clone IMAGE:982855, mRNA
 DEFINITION sequence.
 ACCESSION AA541358
 NID 92287792
 VERSION AA541358.1 GI:2287792
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 314)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1394510.
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D., Rodrigo F. Chuang, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: David B. Kitzman, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL etc.

www.bio.lnl.gov/bbrp/image/image.html
 Insert Length: 389 Std Error: 0.00
 Seq primer: -40m13 fwd. EF from Amersham
 High quality sequence stop: 273.
 FEATURES Location/Qualifiers
 source 1..314
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:982855"
 /clone_lib="NCI_CGAP_Pri2"
 /sex="male"
 /tissue_type="metastatic prostate bone lesion"
 /lab_host="DH10B"
 /note="Vector: PAMP10; mRNA made from metastatic prostate lesion of the bone, cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Kitzman, NIH."
 BASE COUNT 97 a 52 c 65 g 100 t
 ORIGIN
 Query Match 70.3%; Score 24.6; DB 35; Length 314;
 Best Local Similarity 87.1%; Pred. No. 4.8e+02;
 Matches 27; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 Oy 5 CGAGTCGACATCGATTTTCTTTT 35
 ||||| ||||| ||||| ||||| |||||
 Db 313 CGGTCGACTTCATTTTCTTTT 283
 RESULT 10
 A1726205 466 bp mRNA EST 11-JUN-1999
 LOCUS BNGH5252 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (Y08741) beta-tubulin 1 [Hordeum vulgare], mRNA sequence.
 DEFINITION A1726205
 ACCESSION 95045057
 NID A1726205
 VERSION A1726205.1 GI:5045057
 KEYWORDS EST.
 SOURCE upland cotton.
 ORGANISM Gossypium hirsutum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
 REFERENCE 1 (bases 1 to 466)
 Blewitt, M., Matz, E.C., Davy, D.F. and Burr, B.
 ESTs from developing cotton fiber
 JOURNAL Unpublished (1999)
 COMMENT On May 18, 1998 this sequence version replaced gi:3137811.
 CONTACT: Ben Burr
 Biology Department
 Brookhaven National Laboratory
 Upton, NY 11973, USA
 Tel: 516-344-3396
 Fax: 516-344-3407
 Email: burrd@bnl.bnl.gov
 Seq primer: T3 Primer.
 FEATURES Location/Qualifiers
 source 1..466
 /organism="Gossypium hirsutum"
 /cultivar="Acala Maxxa"
 /db_xref="taxon:3635"
 /clone_lib="Six-day Cotton fiber"
 /tissue_type="immature fiber"
 /dev_stage="Six days post anthesis"
 /lab_host="XL1-Blue"
 /note="Vector: pBluescript II KS+"
 BASE COUNT 133 a 118 c 71 g 144 t
 ORIGIN

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNLT at:
www.bio.lnlnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40bp from Glibco
High quality sequence stop: 432.
Location/Qualifiers

FEATURES

source

1. 511
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:915787"
/clone_lib="NCI-CGAP_P12"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

BASE COUNT

143 a 102 c 101 g 165 t

ORIGIN

Query Match 69.7%; Score 24.4; DB 51; Length 511;
Best Local Similarity 82.4%; Pred. No. 4.5e+02;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACTCGAGTCGACATCGATTTTCTTTTCTTTT 35
Db 7 ATTCCGGTCGCCGCGATTTTCTTTTCTTTT 40

RESULT 14

AI734188

LOCUS

DEFINITION

AI734188 517 bp mRNA EST 14-JUN-1999
nc78e12.y5 NCI-CGAP_P12 Homo sapiens cDNA clone IMAGE:783502
similar to contains Alu repetitive element; mRNA sequence.

ACCESSION

AI734188

NID

95055301

VERSION

AI734188.1 GI:5055301

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 517)

AUTHORS

NCI/NIH-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL

On Dec 20, 1995 this sequence version replaced gi:113555.

COMMENT

Other_ESTs: nc78e12.x5

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILNLT at:
www.bio.lnlnl.gov/bbrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40bp from Glibco
High quality sequence stop: 417.
Location/Qualifiers

FEATURES

source

1. 517
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:783502"
/clone_lib="NCI-CGAP_P12"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site 1: NotI; Site 2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the UDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Krizman."

BASE COUNT

129 a 131 c 93 g 164 t

ORIGIN

Query Match 69.7%; Score 24.4; DB 51; Length 517;
Best Local Similarity 82.4%; Pred. No. 4.5e+02;
Matches 28; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 2 ACTCGAGTCGACATCGATTTTCTTTTCTTTT 35
Db 14 ATTCCGGTCGCCGCGATTTTCTTTTCTTTT 47

RESULT 15

AA145612

LOCUS

DEFINITION

AA145612 570 bp mRNA EST 12-FEB-1997
ms07h10.r1 StrataGene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:606307 5' similar to gb:xl2953 RAS-RELATED PROTEIN RAB-2
(HUMAN); mRNA sequence.

ACCESSION

AA145612

NID

91715011

VERSION

AA145612.1 GI:1715011

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 570)

AUTHORS

Marra, M., Kucab, T., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucab, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Thelsting, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE *The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced g1:1292118.

Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.lml.gov) for further information.
MGI:371739
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 347.

FEATURES
SOURCE

1. 570
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone_image="606307"
/clone_id="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1:
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector: ~5'
adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGACGTTTTTTTTTTTTTTT 3' "

BASE COUNT 143 a 115 c 157 g 155 t

ORIGIN

Query Match 69.1%; Score 24.2; DB 29; Length 570;
Best Local Similarity 89.7%; Pred. No. 4.9e+02;
Matches 26; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

UV 7 AGTCGACATCGATTTTTTTTTTTTTTTT 35
|| ||||| ||||| ||||| |||||
Db 314 AGTCGACATCGATTTTTTTTTTTTTTTT 342

Search completed: September 13, 1999, 15:45:55
Job time: 4043 sec

Query Match 100.0%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0046;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTAAAGCTTAAGTCAATCATCC 35
|||||

DB 1 TTAAAGCTTAAGTCAATCATCC 35

RESULT 2
BOVPHOS3K/c
LOCUS Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds.
DEFINITION
ACCESSION M93252
VERSION M93252.1 GI:163519
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE Bos taurus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinn, S., Fry, M.J., Gout, I., Dhan, R., Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D. Phosphatidylinositol 3-kinase: Structure and expression of the 110 kd catalytic subunit
JOURNAL Cell 70, 419-429 (1992)

FEATURES
source location/Qualifiers
1..3207
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SC6AR-1"
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AAA30698.1"
/db_xref="PIR:G163520"
/db_xref="GI:163520"

CDS
1..3207
/translacion="MPPRSSEGLMGHILMPRIIVECLIPNGMIVTECLREALTIT
IKHELFEARKYPLHQLDDESSYIFSVTOEAREEPEFDETRLCIDLRLFPPLKVI
EFGVREBEKILNREIGFALGMVCEDEVDVFRNINLWCKEAVDLNLSPH
SRAMVYPPNVESSPELPHKYNLKDGOIIVIVWVSPNDKOKYTLKINDCVPE
OYIAEAIKRTSRMLSSBOLKLVLEOGKYLIVCGDEVEYELKPLSOVKYTRSC
IMLGPMNMLMAKESLYSQLPMDCTFMRSTSRISTATPYMNGETSTKSLMINSAL
RIKICATVYVNNIRIDIDKIVRTGIYHGEPLCDNVTPVPCSNPRNEMLVNDIY
IPDLPRARLCLISICSVKGRKAKHECHPLANGNINLEDYDTLVSGLMALNMPVH
GLEDLNPIGVTSNPNKETPELLEDMDVSVKPFDMVSYEEHANNVSREAGFSY
SHAGLSNRILARDNELRENDKQRLAICTRDPLETTEDKDFLMSHRHVCYIIPILP
KLISVKNRSDVEVAOMYCIYDMPKIDPEOMELLDYCNYPWRGFAVRLLEYITL
DDKLSQYLIQVLYKTEQYLDNLVRLFKALNORIGHEFFHKLSEHANKTVSQ
RFGDLSEYCRACGMYLKHLNRYEAMEKLINLDTLQEKEDTOKVOMKFLVEONR
RPFDMALOGFSLNPAHQNLNRECRIMSARKPLMLNMPNEDIMSEILFQNR
IIFKNGDLRQMDLTQIIRIMENIMNOGDLRMLPYGCLISGCVGLIEVRSHT
IMQIOCGGLKALQFNSHTLHOMLKNNKEIYDAIDLFTRSCAGCVATFTLIGIG
DRHNSNIVKWDGOLFHTDEGHFLDKKKKREKREKRPVLTQPLVVISGAGBECT
KTRFEROEKCYATVLAIRQHANLFINLSMGLSGMPELQSFDDIAYIRKTLALDK
TEQALLEYFMQMDAHHGWTGTMDFHTIKOHALN"

BASE COUNT 1028 a 581 c 680 g 918 t

ORIGIN

Query Match 74.3%; Score 26; DB 3; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AGGCAATTCATTAAGTCAATCATCC 35
|||||

DB 10 AGGCAATTCATTAAGTCAATCATCC 35

DB 110 AGGCAATTCATTAAGTCAATCATCC 85
|||||

RESULT 3
A37232/c
LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
VERSION 92294345
NID 92294345
KEYWORDS A37232.1 GI:2294345
SOURCE unidentified.
ORGANISM unidentified.
COMMENT unclassified.
REFERENCE 1 (bases 1 to 3498)
Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D. EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL IMP CANCER RES TECH (GB)
Other publication JP 8503124T 960409.
COMMENT Location/Qualifiers
1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PIR:e304808"
/db_xref="PIR:92294346"
/db_xref="GI:2294346"

CDS
1..3207
/translacion="MPPRSSEGLMGHILMPRIIVECLIPNGMIVTECLREALTIT
IKHELFEARKYPLHQLDDESSYIFSVTOEAREEPEFDETRLCIDLRLFPPLKVI
EFGVREBEKILNREIGFALGMVCEDEVDVFRNINLWCKEAVDLNLSPH
SRAMVYPPNVESSPELPHKYNLKDGOIIVIVWVSPNDKOKYTLKINDCVPE
OYIAEAIKRTSRMLSSBOLKLVLEOGKYLIVCGDEVEYELKPLSOVKYTRSC
IMLGPMNMLMAKESLYSQLPMDCTFMRSTSRISTATPYMNGETSTKSLMINSAL
RIKICATVYVNNIRIDIDKIVRTGIYHGEPLCDNVTPVPCSNPRNEMLVNDIY
IPDLPRARLCLISICSVKGRKAKHECHPLANGNINLEDYDTLVSGLMALNMPVH
GLEDLNPIGVTSNPNKETPELLEDMDVSVKPFDMVSYEEHANNVSREAGFSY
SHAGLSNRILARDNELRENDKQRLAICTRDPLETTEDKDFLMSHRHVCYIIPILP
KLISVKNRSDVEVAOMYCIYDMPKIDPEOMELLDYCNYPWRGFAVRLLEYITL
DDKLSQYLIQVLYKTEQYLDNLVRLFKALNORIGHEFFHKLSEHANKTVSQ
RFGDLSEYCRACGMYLKHLNRYEAMEKLINLDTLQEKEDTOKVOMKFLVEONR
RPFDMALOGFSLNPAHQNLNRECRIMSARKPLMLNMPNEDIMSEILFQNR
IIFKNGDLRQMDLTQIIRIMENIMNOGDLRMLPYGCLISGCVGLIEVRSHT
IMQIOCGGLKALQFNSHTLHOMLKNNKEIYDAIDLFTRSCAGCVATFTLIGIG
DRHNSNIVKWDGOLFHTDEGHFLDKKKKREKREKRPVLTQPLVVISGAGBECT
KTRFEROEKCYATVLAIRQHANLFINLSMGLSGMPELQSFDDIAYIRKTLALDK
TEQALLEYFMQMDAHHGWTGTMDFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t

ORIGIN

Query Match 74.3%; Score 26; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AGGCAATTCATTAAGTCAATCATCC 35
|||||

DB 110 AGGCAATTCATTAAGTCAATCATCC 85

RESULT 4
HSPH3K/c
LOCUS HSPH3K 3424 bp mRNA PRI 24-AUG-1995
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION Z29090
VERSION 9472990
NID Z29090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 1068)
AUTHORS Volinia, S., Hiles, I., Ormondroyd, E., Nizetic, D., Antonacci, R., Rocchi, M. and Waterfield, M.D.
TITLE Molecular cloning, cDNA sequence, and chromosomal localization of the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL Genomics 24 (3), 472-477 (1994)
MEDLINE 95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia, S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig Institute for Cancer Research, 91 Riding House Street, London, W1P 8BT, UK

FEATURES
source Location/Qualifiers
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="KGla"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CA82333.1"
/db_xref="PIR:G472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPPRSSGSELMGHIIMPRLIVBCLLPNGMIVTECLREATLV
IKHLEFKARKYPIHQLDDESSYIFVSVDAREEFEDTRICDLRLQPLKVI
EPVNNREKILNREIGALIGPVEFDWVQDPRNRLINVCKEAVDRDLNSH
SRNAVTPPVNESSPELPHKILNLDGQILVIVLVSPNNDOKYTLKINHCVBE
QVIAEAIKTRKSMILSSEOLKLCVLEOYGYILKVCOCDEPLEKYLISOYKIRSC
IMLGRMPLMLMAKESLYSOLPMDCFMPSYSRISATPYMNGESTSKSLAVINSL
RIKILCAVYVNRIDIKIVRTGIHGSEPLCDNVTQVPCSNRMNMLNVDYI
IPDIPRARILCLSCYKGRKGAEEHCPPLAMGNINLEFDYDITLVSGMALNMPVH
GLEDLNPIGVTSNPKETPCLELEDFWESSYVKFPDMSVIEEHANWSREGEFSY
SHAGLSNRLANDNELRENDKQLAISTRDLSITDEKDFLMSHRYCVTIEILP
KLLLSVWNSRDEVAOMYCLVKNPPIKEQAMELLDCNPDPMVRGFAVCLERYL
DKLSQYLIOVLYKYEQYLDNLVFLKALTNORIGHFFMHLKSEHNVTVSQ
RFGLLSYCACGMVYLKHLNROYEAMEKILINDIIOKEKDETOYOMKFIYVQNR
RPDMDALOGFLSPINPAHOIGNRLEKCRIMSSAKRPLIMENPDIMSLLQNNR
IIFKNGDLRODMITLTIIRLMENIMONGIDLMLPYGCLSIDCVGLIIVYNSHT
IMQIOCGGLGALQFNSHTLHOMLKDNKKEIYDAIDLFTRSCAGCYATFLLIG
DRHNSINIKDQGLFHDGFHDLHKKKRGYRERPEVLTQDFLIVISKGOECT
KTRREFEOMCYAAYLAIROHANLFIINLSMMLGSGMPELQSFDDIAYIKTIALDK
TEQDALEFPMQMDAHNGWTTKMDVFTIKOHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 69.7%; Score 24.4; DB 10; Length 3424;
Best Local Similarity 96.2%; Pred. No. 2.6;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCACTCTAAAGTCATCATCC 35
|||||
Db 122 AGGCACTCTAAAGTCATCATCC 97

RESULT 5
LOCUS HS079143/c 3207 bp mRNA PRI 02-JAN-1997
DEFINITION Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA,
complete cds.
ACCESSION U79143
VERSION U79143.1 GI:1763625
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3207)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE Catalytic Activity of the p110-alpha Subunit of Human Phosphoinositide 3'-Hydroxykinase is Required for Signal Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Sturdivant, S.M., Ahern, J., Conroy, R.R., Barnett, S.F., Ledder, L.M., Oliff, A. and Heimbrock, D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs, Sunnytown Pike, West Point, PA 19486, USA

FEATURES
source Location/Qualifiers
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/product="phosphoinositide 3'-hydroxykinase p110-alpha subunit"
/codon_start=1
/protein_id="AAB39753.1"
/db_xref="PIR:G1763626"
/db_xref="GI:1763626"
/translation="MPPRSSGSELMGHIIMPRLIVBCLLPNGMIVTECLREATLV
IKHLEFKARKYPIHQLDDESSYIFVSVDAREEFEDTRICDLRLQPLKVI
EPVNNREKILNREIGALIGPVEFDWVQDPRNRLINVCKEAVDRDLNSH
SRNAVTPPVNESSPELPHKILNLDGQILVIVLVSPNNDOKYTLKINHCVBE
QVIAEAIKTRKSMILSSEOLKLCVLEOYGYILKVCOCDEPLEKYLISOYKIRSC
IMLGRMPLMLMAKESLYSOLPMDCFMPSYSRISATPYMNGESTSKSLAVINSL
RIKILCAVYVNRIDIKIVRTGIHGSEPLCDNVTQVPCSNRMNMLNVDYI
IPDIPRARILCLSCYKGRKGAEEHCPPLAMGNINLEFDYDITLVSGMALNMPVH
GLEDLNPIGVTSNPKETPCLELEDFWESSYVKFPDMSVIEEHANWSREGEFSY
SHAGLSNRLANDNELRENDKQLAISTRDLSITDEKDFLMSHRYCVTIEILP
KLLLSVWNSRDEVAOMYCLVKNPPIKEQAMELLDCNPDPMVRGFAVCLERYL
DKLSQYLIOVLYKYEQYLDNLVFLKALTNORIGHFFMHLKSEHNVTVSQ
RFGLLSYCACGMVYLKHLNROYEAMEKILINDIIOKEKDETOYOMKFIYVQNR
RPDMDALOGFLSPINPAHOIGNRLEKCRIMSSAKRPLIMENPDIMSLLQNNR
IIFKNGDLRODMITLTIIRLMENIMONGIDLMLPYGCLSIDCVGLIIVYNSHT
IMQIOCGGLGALQFNSHTLHOMLKDNKKEIYDAIDLFTRSCAGCYATFLLIG
DRHNSINIKDQGLFHDGFHDLHKKKRGYRERPEVLTQDFLIVISKGOECT
KTRREFEOMCYAAYLAIROHANLFIINLSMMLGSGMPELQSFDDIAYIKTIALDK
TEQDALEFPMQMDAHNGWTTKMDVFTIKOHALN"

BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN

Query Match 69.7%; Score 24.4; DB 10; Length 3207;
Best Local Similarity 96.2%; Pred. No. 2.6;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCACTCTAAAGTCATCATCC 35
|||||
Db 110 AGGCACTCTAAAGTCATCATCC 85

RESULT 6
LOCUS MM003279/c 3207 bp mRNA ROD 25-MAY-1994
DEFINITION Mus musculus Balb/c phosphatidylinositol 3-kinase p110 kDa subunit
mRNA, complete cds.
ACCESSION U03279
VERSION U03279.1 GI:414994
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3207)

AUTHORS Klippel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
TITLE The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE 94187738
REFERENCE 2 (bases 1 to 3207)
AUTHORS Klippel-Giese, A.
TITLE Direct Subunit
JOURNAL Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dailchi Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
LOCATION/Qualifiers 1..3207
FEATURES /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"
 1..3207
CDS /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA1834.1"
 /db_xref="pid:9414995"
 /db_xref="gi:414995"
 /translation="MPPRSSGELMGTHLMPRIIVVCELLPNGMTVTECLREATLYT
 IKHELFRARKYPLQLQDSSITLVSTOAEEREFDETRRLCDLRLPPLKVI
 EPGVREKRLREIGFAIGMPEFNDKDEYODFRNTLNCKEANDLDNAPH
 SPALVYPPNVSSELPKRIHYNKLDKQIIVYIVVSPNNDKQYTLKINHCVPE
 QVIAIRKTRSMLSSEQLKLVLEYQKYLIVVCGDEYFLERKYSQYIRSC
 IMLGMPMLMAKESLYQLIDSTFMPYSRISTATPYNGESTLSQYIRSC
 RLKICATVYVNIIDRIKYRTGTYHGGEPLCDNVTORVCSNPRNEMLVNDAI
 IPDLRLARLCLSCSVKGRGAKGHEHPLANGINLIPDTYTLVSGKMAILMPVH
 GLEDLNPGVTSNPKETPCLEPDMFSVYKPPDKSVIEEHNVSVEAGFSY
 SHTGISNRLARNDRENDKQLALCTRPDLSSETDEKDLMSHRHCVIPILP
 KLLSVKWNRSDEVAQMYCLVDPWPIKPEQAMELDCYPPPWRSFAVCLLEYLT
 DDKLSQYLIQVQVLYKEQYLDNLVREFLKRALTNORIGHEFFWHLKSEMNKTVSQ
 RFGLLSEYCAACGWTYKHLNROVAMEKILNLTDLKOEKDETOYKMFLEVOMR
 OPDFMDALOGELSPINPAHQNLNLEECRIMSARKRPLMWENPDIMSELFPONE
 IIFKNGDRLQDMTLQIIRIMENIWNQGLDRLPYGCLSIGCVGILEYRNSHT
 IMQIKCGKGLAQFNSTHLQWLKDNKKGITPAIDLFTRSCAGTCVATFIIIGT
 DRHNSNINWKDQGLFHDGFHDLHKKKFGYKREVEFVLTQDFLIVISGAOECT
 KTRFERFOEMCYKAYALIRQHANLFINFSMLGSGMBELQSTFDIAYIRKTLADK
 TEQEALEYFMQMDAHHGWTTKMDWIFTHIKOHALN"

BASE COUNT 958 a 693 c 734 g 822 t
ORIGIN

Query Match 69.7%; Score 24.4; DB 12; Length 3207;
Best Local Similarity 96.2%; Pred. No. 2.6;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Cy 10 AGCATTCCTAAGTCACTATCATCC 35
 |||||||||
 |||||||||

Db 110 AGCATTCCTAAGTCACTATCATCC 85

RESULT 7
LOCUS AF001076 3452 bp mRNA VRT 08-JUL-1997
DEFINITION Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
 complete cds.
ACCESSION AF001076
NID 92245505
VERSION AF001076.1 GI:2245505
KEYWORDS chicken.
SOURCE Gallus gallus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3452)
AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase

JOURNAL Science 276 (5320), 1848-1850 (1997)
MEDLINE 97354438
REFERENCE 2 (bases 1 to 3452)
AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
TITLE Direct Subunit
JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
LOCATION/Qualifiers 1..3452
FEATURES /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /tissue_type="brain"
 1..150
 151..3354
 /gene="c-p3k"
 /note="p3k proto-oncogene"
 151..3357
 /note="c-p3k"
 /codon_start=1
 /product="phosphoinositide 3-kinase catalytic subunit"
 /protein_id="AAB62534.1"
 /db_xref="pid:9245506"
 /db_xref="gi:2245506"
 /translation="MPPRSSGELMGTHLMPRIIVVCELLPNGMTVTECLREATLYT
 IKHELFRARKYPLQLQDSSITLVSTOAEEREFDETRRLCDLRLPPLKVI
 EPGVREKRLREIGFAIGMPEFNDKDEYODFRNTLNCKEANDLDNAPH
 SPALVYPPNVSSELPKRIHYNKLDKQIIVYIVVSPNNDKQYTLKINHCVPE
 QVIAIRKTRSMLSSEQLKLVLEYQKYLIVVCGDEYFLERKYSQYIRSC
 IMLGMPMLMAKESLYQLIDSTFMPYSRISTATPYNGESTLSQYIRSC
 RLKICATVYVNIIDRIKYRTGTYHGGEPLCDNVTORVCSNPRNEMLVNDAI
 IPDLRLARLCLSCSVKGRGAKGHEHPLANGINLIPDTYTLVSGKMAILMPVH
 GLEDLNPGVTSNPKETPCLEPDMFSVYKPPDKSVIEEHNVSVEAGFSY
 SHTGISNRLARNDRENDKQLALCTRPDLSSETDEKDLMSHRHCVIPILP
 KLLSVKWNRSDEVAQMYCLVDPWPIKPEQAMELDCYPPPWRSFAVCLLEYLT
 DDKLSQYLIQVQVLYKEQYLDNLVREFLKRALTNORIGHEFFWHLKSEMNKTVSQ
 RFGLLSEYCAACGWTYKHLNROVAMEKILNLTDLKOEKDETOYKMFLEVOMR
 OPDFMDALOGELSPINPAHQNLNLEECRIMSARKRPLMWENPDIMSELFPONE
 IIFKNGDRLQDMTLQIIRIMENIWNQGLDRLPYGCLSIGCVGILEYRNSHT
 IMQIKCGKGLAQFNSTHLQWLKDNKKGITPAIDLFTRSCAGTCVATFIIIGT
 DRHNSNINWKDQGLFHDGFHDLHKKKFGYKREVEFVLTQDFLIVISGAOECT
 KTRFERFOEMCYKAYALIRQHANLFINFSMLGSGMBELQSTFDIAYIRKTLADK
 TEQEALEYFMQMDAHHGWTTKMDWIFTHIKOHALN"

BASE COUNT 3'UTR 1073 a 672 c 783 g 924 t
ORIGIN

Query Match 65.1%; Score 22.8; DB 4; Length 3452;
Best Local Similarity 92.3%; Pred. No. 11;
Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Cy 10 AGCATTCCTAAGTCACTATCATCC 35
 |||||||||
 |||||||||

Db 260 AGCATTCCTAAGTCACTATCATCC 235

RESULT 8
LOCUS AF001075 3389 bp mRNA VRT 08-JUL-1997
DEFINITION Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
 subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
 partial cds.
ACCESSION AF001075
NID 92245502
VERSION AF001075.1 GI:2245502
KEYWORDS avian sarcoma virus 16.
SOURCE Avian sarcoma virus 16
ORGANISM Viruses; Retroviridae; Retrovirinae; Avian type C retroviruses.
REFERENCE 1 (bases 1 to 3389)
AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.

TITLE Transformation of chicken cells by the gene encoding the catalytic subunit of p13-kinase

JOURNAL Science 276 (5320), 1848-1850 (1997)

LOCUS 97334438

REFERENCE 2 (bases 1 to 3389)

AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.

TITLE Direct Submission

JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop BCC-239, La Jolla, CA 92037, USA

FEATURES

source 1..3389

organism="Avian sarcoma virus 16"

db_xref="taxon:60629"

gene <1..33252

gene="gag-v-p3k"

/note="fusion gene of gag and v-p3k"

1..78

gene="gag-v-p3k"

/note="derived from gag gene"

<1..33252

gene="gag-v-p3k"

/codon_start=1

/product="gag-v-phosphoinositide 3-kinase catalytic subunit fusion protein"

/protein_id="AAB62532.1"

/db_xref="PID:g2245503"

/db_xref="GI:2245503"

/translation="QQLIRAPSTLTTPGELIKYVLDKQKSHLMPRLVECLPNG MIVLECEATLTLTIKHELEKARKYPLVOLLDESSYIVSVTOAEKEFEEDTR RICDLRFOPRIKYIEPVNGEELIKRKIFALGPRICEPDMDKDEVDYDFRRITL VCKRAVLDKRNASHALYCPNPVSSPELPHIINKDKGQITVAVIVSPND KQKTLINDCVEQVYIAEIRKTSMLSSQLKCVLEYGKYLKVCQDEYL LKQPLSYKIRSCIMGRPNMLMAKESLYTLPDFTMPSYSKRISTAPYM GEATKSLMTINSLRIRICATVANNIRIDIKYVGTGHGEPICDVNORVP CSNRMNEMSYDMYIPDLPRARLCLSCVGRKGAKEEGCLAMGNIMEYDTT LVSQKMLNMAVPHGLDILNIPGVGNSNKRKTCLELEDFMPSVPRPDVTIE EHAMWTSRELGFTYAGUSNRIRARDNELRESREQLRAICTDPSLSETEQKDFL WSHRYCVKTEPELLPKLLSVKNSRDEAVMYLVDMPIKEQAMELLDCYPPD WYRAFAVCEKLYLDKLSQYLQVLVLEBYLDLQVRLPKLAKLTORGHF FHLKSEMHNTVQORGLLESYCRAGMGLKSLROVEMERLINLTOLKQEKD ETQVOKKFLVEQMRPDMALOGFISPLPAHOLGLREGRIMSSARPMIANK ENDMISELLFONNELIFKNGDILRODMTLQIIRIMENTONGDLRLMPTCLSL GDCVGLLEVRSSTIMQIOCKGGLKGLQFNSTLRQWLNKNGEITDAIDLETR SCAGYCVATFLLGDRHNSNINWKGQLEFHDGFLDKKKFGKRRVPEVLT QDLDIVISKGOECTKREFEFQEMCHKAYLAIRHANLEINFSMGLSGMPELDS FDIAYIRKTLALDKTEQEALEYFMKNDAHGGWTKMDWIFHTIKOHALN"

misc_feature 79..84

gene="gag-v-p3k"

/note="sequence of unknown origin"

85..33252

gene="gag-v-p3k"

/note="derived from c-p3k gene"

3250..3370

CDS 3371..>3389

/codon_start=1

/product="retroviral env"

/protein_id="AAB62533.1"

/db_xref="PID:g2245504"

/db_xref="GI:2245504"

/translation="MARNS"

BASE COUNT 1071 a 655 c 751 g 912 t

ORIGIN

Query Match 65.1%; Score 22.8; DB 17; Length 3389;

Best Local Similarity 92.3%; Pred. No. 11;

Matches 24; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 AGGATCTAAAGTACTATCATCC 35

DB 155 AGGATCTAGAGTACATCATTC 130

RESULT 9

CET04A8/C

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

JOURNAL

TITLE

REFERENCE

AUTHORS

JOURNAL

COMMENT

TITLE

JOURNAL

DEFINITION

COMMENT

94150718

Coding sequences below are predicted from computer analysis, using predictions from GeneFinder (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: -

http://webc.sanger.ac.uk/cgi-bin/display?db=wormacc&class=sequence&object=T04A8

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note. IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring subclones. IMPORTANT: This sequence is not the entire insert of clone T04A8. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring subclones. The true left end of clone T04A8 is at 1 in this sequence. The true right end of clone T04A8 is at 4879 in sequence Z35719. The true right end of clone T23f11 is at 5962 in this sequence. The start of this sequence (1..96) overlaps with the end of sequence Z46343. The end of this sequence (39470..39482) overlaps with the start of sequence Z35719. Location/Qualifiers

1..39482

/organism="Caenorhabditis elegans"

/db_xref="taxon:6239"

/chromosome="III"

/clone="T04A8"

```

gene      285..2182
CDS       /gene="arg-10"
          /join(285..515,1092..1310,1365..1508,1567..1749,1985..2182)
          /note="similar to 7tm receptor"
          /codon_start=1
          /protein_id="CA84725.1"
          /db_xref="PID:e1348902"
          /db_xref="PID:g3879340"
          /db_xref="GI:3879340"
          /db_xref="SWISS-PROT:P46568"
          /translation="MNSSRVPPANFSEYEDPLPEECNEDPNVVLISAMYGMSYILI
          GAVLIMVITYEFGHSYRDNFMYLQADATVGYITNTAEYFGRIETITLPIPIA
          SPYFETPSILTKMYNAALHSIGFKTFISQIFNSFNMTVITLAKHLKIKQILKPLV
          IITFLPLGVNKLILSRVYINNGAGFVSNKDFPANNISILHLEHETLCEVAVII
          FVATVILGLTLMKORIKSAERSLITVTMAVQVAFASIQIYFVEFAVYTPKIRSVL
          LOIVSEVFDLSLVEFSPIALIVMSRQIRKIDIFMLKDETQISMYPNSEL"
          2532..3697
gene      /gene="arg-11"
CDS       /join(2532..2658,2878..3338,3506..3697)
          /gene="arg-11"
          /note="See T23F11.5"
          /codon_start=1
          /protein_id="CA84735.1"
          /db_xref="PID:e1348912"
          /db_xref="PID:g3879350"
          /db_xref="GI:3879350"
          /db_xref="SWISS-PROT:P46569"
          /translation="MPSRPNISNDITIIIFECNSNYDITVEYKMFLOIAYLIPGI
          LNILLYTLIFKNSIEYASSFEFLYSDCFSESMIFDIIGRTLYTTPCTPIAP
          MEYEPIDGKIMVILHNSRACKSLIQLILVYRMSCVIYPIRYGKMMRPKLYIIL
          VVIEPESIDNMLIISRYVMOQPEGGIYMEYIKVMASSRQLFITALLFTYVCT
          SVIFTVLVPRLRNVERKISLGLTNTISMSTIILVFOVN"
          complement(3293..7300)
gene      /gene="T04A8.3"
CDS       /complement(join(5293..5696,5899..6060,6234..6659,
          6706..6965,7019..7072,7227..7300))
          /gene="T04A8.3"
          /note="similar to C-type lectin"
          /codon_start=1
          /protein_id="CA84726.1"
          /db_xref="PID:e1348903"
          /db_xref="PID:g3879341"
          /db_xref="GI:3879341"
          /db_xref="SPTREMBL:Q22136"
          /translation="MFDVTIVRTMLSHDIFKMRNSTRMKVTNARSEFLKRTMRN
          PLSTCEDNVPYHINNAESEYKTCCEATLICSQVOMFGRCYKITKMTEDKA
          BEHCKNOQDHTSTIAFMHREALPFMNDYFTVSRIMQDASKVITINDLIYDEGVNL
          LALDGYKYLPNVAIARVPKDELTANVLCEYTPPKTKSESNTLLRRYGEIYYPLVTS
          SVYMTTSSRIRINADPLADNVCYELMKPVRGCEAQSALPTQEFYKKTJCTNFRF
          YRKYTGOLMKCKIYHDSMNNDNAEKQNDGAHLSFTTIEELKLDDEMLLEYPN
          DNIIAVWLGAKRRECGDLSKNFTGYSKDIDHPARSVFEMQNGVANOPIFVGDG
          FDYMAKEYEPNSTDSERCLYQMSGSMVYGDNPRNMQINDIYCNTEFFKLCGEA
          TFFAVKKNL"
          complement(7871..9714)
gene      /gene="T04A8.4"
CDS       /complement(join(7871..8071,8723..8819,8916..9181,
          9228..9359,9489..9585,9638..9714))
          /gene="T04A8.4"
          /codon_start=1
          /protein_id="CA84722.1"
          /db_xref="PID:e1348898"
          /db_xref="PID:g3879337"
          /db_xref="GI:3879337"
          /db_xref="SPTREMBL:Q22133"
          /translation="MPISASDTIRKFNAAVANOANEVLKKNPYSDTYIKQAFDTKNG
          RPPGSKTEARIGIKAGVHCRIELFLCEIDSNADGEPHKYKFGKLEPNISYFSDK
          LVGMLIRAKYGLVHEGEMLYARODDEKITIMMALSDEIRSKVILASGDPRANCIOIR
          NSEPAVYPPVSKKAPRPPIRSDSVTSKAFKFAAPPNDISLTVQGPKNKPTWPKD
          TSKAPILIRRRRLKIFMGLCLCYTNAQVDHDSFQOMCEFMGGEQTYRANGYKVO
          IYDLSIDIMDLN"
          11864..14597
gene      /gene="T04A8.5"

```

```

CDS       /join(11864..11952,12576..12981,13028..13288,13524..13833,
          13882..13997,14293..14457,14502..14597)
          /gene="T04A8.5"
          /note="Amidophosphoribosyltransferase (glutamine
          phosphoribosylpyrophosphate amidotransferase)"
          /codon_start=1
          /protein_id="CA84723.1"
          /db_xref="PID:e1348900"
          /db_xref="PID:g3879338"
          /db_xref="GI:3879338"
          /db_xref="SPTREMBL:Q22134"
          /translation="MCGIFGVANGVNLVNLVANGALAOHRESESTGVGSDGIT
          RDHETILGCHLVADYITEDNISRNQSGIITIHNRKSTACKKSGICVQPPVYTA
          MGVAIANGELVDAKQKREVLHGEVLYSTDUSELTAQMIARAIALNVRCKYQEM
          GDITRELVATSNMYSLSLVMTFEDRLYLRADFGNRLCVGVYSKNGPEAFIAS
          SESCFFPNAKLDEFEVRGEIYELSTGKIRVOMKPTPLAMCFEYVYPRANDSEI
          EGOQVQYREEGCTMALEDLEADIVNGVDSLSAIGVASOGLIYEVLRNSY
          VGRSPTEPDEMROMNATKMRGVYKKYKHGQRTVLYNDSIYRGNTMRTVYMLDACA
          KEVHLRIASPPVKFPCFNGINPTTKELIAAEKTIPEICQFVGADSVRYLSVDGVSS
          VQKIERANFSPGHCCTACTGKTPVALDL"
          complement(14747..15807)
gene      /gene="T04A8.6"
CDS       /complement(join(14747..14836,14881..15396,15445..15558,
          15604..15807))
          /gene="T04A8.6"
          /note="similar to ribonucleoprotein; cDNA EST yk222a11.3
          comes from this gene; cDNA EST yk432f10.3 comes from this
          gene; cDNA EST yk432f10.5 comes from this gene; cDNA EST
          yk432f10.5 comes from this gene; cDNA EST yk497a8.3 comes
          from this gene; cDNA EST yk497a8.5 comes from this gene"
          /codon_start=1
          /protein_id="CA84724.1"
          /db_xref="PID:e1348901"
          /db_xref="PID:g3879339"
          /db_xref="GI:3879339"
          /db_xref="SPTREMBL:Q22135"
          /translation="MVAONPSSISGGGKOPKODKYOKOLKIYVYKIRIPFGEF
          EKELGYRQGNVLRIRVANNRRPTGKNGKAYIGFDKAAVAELTAASMGYLFGEOR
          LGCTVMKPELIPKARHGPFLVMPSPYIGAKKPTIARNNTGNDVIYNAKRYOAN
          KTLRKLQMGINDYFVSQGTPTPLKVIDEVOFISDVPKSTPIIVKAAATPKAAT
          PKAATPKAATPKVSTPYTKVYIKTPGKTEKPEQPTPKTBAGKAAAAATPAKSTL
          SNTILKSYAASARPAVAEKKTLRSRKKKS"
          16523..17601
gene      /gene="T04A8.8"
CDS       /join(16523..16784,16831..17000,17047..17601)
          /gene="T04A8.8"
          /note="cDNA EST yk339d3.3 comes from this gene; cDNA EST
          yk339d3.5 comes from this gene"
          /codon_start=1
          /protein_id="CA84734.1"
          /db_xref="PID:e1348911"
          /db_xref="PID:g3879349"
          /db_xref="GI:3879349"
          /db_xref="SPTREMBL:Q22144"
          /translation="MDSDDNAYSDIYGNRAALQKKRPYRVCSDDSDDEPTSGP
          SPARKTRETQLPDDSDDECAVEKEVETISQKRRHLIDYDEVAIRSKIRRETVIA
          PVERNNVITSGLKICEGWDENAKGGAQPEVEDYVDDDADESSIAAENFPVTVIDLC
          65.1%; Score 22.8; DB 36; Length 39482;
          Best Local Similarity 79.4%; Pred. No. 6.4;
          Matches 27; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

QY      1 TTTAAGCTTAGGCATCTAAGTCACATCATCC 34
       111111111111111111111111111111
DB      4747 TTTCAGAAATTGGCATTCTAAGTCATAATTAAAC 4714
       111111111111111111111111111111
RESUTL 10
AF076243 AF076243 99384 bp DNA PLN 11-MAY-1999
LOCUS Arabidopsis thaliana BAC T26N6 from chromosome IV at 19.3 CM,
DEFINITION complete sequence.
ACCESSION AF076243

```

NID 93309276
 :TAXON AF076243.1 GI:3309276
 KEYWORDS HTG.
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Magnoliophyta; Magnoliopsida; core eudicotyledons; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1 (bases 1 to 99384)
 AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
 TITLE Genomic sequence of Arabidopsis thaliana BAC T26N6, chromosome IV, 19.3 cM
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 99384)
 AUTHORS Zhong,J., Ma,P., Parnell,L.D., Chen,C.-N. and Chen,E.Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUL-1998) Applied Biosystems, Division of Perkin Elmer, 850 Lincoln Center Drive, Foster City, CA 94044, USA
 REFERENCE 3 (bases 1 to 99384)
 AUTHORS Parnell,L.D.
 TITLE Direct Submission
 JOURNAL Submitted (11-MAY-1999) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA
 REMARK Arabidopsis thaliana BAC T26N6 from chromosome IV at 19.3 cM
 COMMENT BAC T26N6 was sequenced as part of the Arabidopsis genome sequencing effort of the Cold Spring Harbor Consortium. For additional information, please see <http://www.cshl.org/Arabidopsis>.
 FEATURES
 source Location/Qualifiers
 1..99384
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="IV"
 /map="19.3 cM"
 /clone="BAC T26N6"
 /note="hybridizing YAC is C1C8B1"
 misc_feature 1..9514
 /note="Overlap with T19B17, GenBank accession number AF069441, from position 1 to 9513; there is a 1-bp discrepancy between T26N6 and T19B17"
 repeat_region 4268..4587
 /note="function-unclassified; similar to K23L20, GenBank accession number AB016874; similar to T19K24, GenBank accession number AC002342"
 gene 5477..6145
 /rpt_type="dispersed"
 /gene="T26N6.1"
 /note="similar to histone H2A; the potential protein has a premature COOH terminus, an extremely low similarity to other H2A histones at the amino terminus, and similarity to histones on both strands of T26N6"
 /pseudo
 repeat_region 10402..10406
 /note="function-putative-target duplication site; cognate A of target duplication site flanking retrotransposon from position 10402 to 15747; cognate B resides from position 15743 to 15747"
 repeat_region 10402..15747
 /rpt_type="flanking"
 /note="function-putative-retrotransposon"
 LTR 10407..10796
 /rpt_type="dispersed"
 /note="cognate A of LTR flanking retrotransposon from position 10402 to 15747; cognate B resides from position 15353 to 15742"
 gene 10943..14508
 /gene="T26N6.2"
 /note="encodes putative polypeptide; gene model last

CDS
 edited on 4 May 99"
 /evidence="not_experimental"
 join(10943..12205,12445..12933,13131..13410,13487..14508)
 /gene="T26N6.2"
 /note="functional catalog ID=90"
 /codon_start=1
 /evidence="not_experimental"
 /product="putative polypeptide"
 /protein_id="AAD29768.1"
 /db_xref="PDB:4773895"
 /db_xref="GI:4773895"
 /translation="MENPKRAITSVLKGNYLLAMRTKYVLCGRGLMSYVERKHV
 FKATKEDVDSGFKREEAELEKAKFQEDQNALIQHLESSEILEAVSCETARE
 LMETLENYGVNMTVFEYVKAINNLSQVDEFTKFGFRLMAELMLRSTVD
 PALINERKDEQFGLTLTPANFNDLKHLLRADKLPSLENVCSQVQKEGSLGS
 GKGLITAHKQVIGKGRKVVVCHCKGKMKVCKVHLPAKFAKIGSOEVS
 DQGEVVRKSLDILRSIASIKESGTFLEPKMKLKESTSEFTSEPSSTLIDSG
 ASHMINNPSLIDIKRALGVAVANDKPYVKEIGELNLFDKSKALYMSFTSNLL
 SVKRAIDNLNCTYIFGPNVHFQDKIGRSLLTYENCFDLVHSVWSPCKSRNKT
 FVTEIDERSKTYITLLPSKCDVADAEINQNYVTNFMKIKVLRDNGEYTSK
 KEYLAKGIIHQISCPYTPQONGVAEERKRLMEVSRMFTVTKFMDAVMTAC
 YLIRTPTKVQDVSTFEFEKGYEKDQNSLADSTPIDRATSLQFLDLGVT
 PSREKTRDLIEPITIDQENQFASNLQDGEINGIQIHDMQIQNDQGEVIG
 IREKSSRLTYNNKAVAPIQAVCSLALPDQHOAFIKTANFVLQYTEKSEBETI
 NAVADETAMTRHFWDEEDLPGRKAVSSWVFTITLSGELERHAKRLVACFTQ
 TYGRDYETETAPVAKLHTRVRLVSLATNLSDWLMQVKNKFLQGELEEEVYMPPG
 LEDSIAGKYLRLKALTYGLKOSPRVHNSITLKGKFSADHTLFLQSDGI
 IVALIYVDNIISGDNKEGIDHTRFLKFSKFDIDGLKFLIEVCRSPEGFLSQ
 RKYTLDLNENGKIGSKPAKPLVDGYTKRTGGRRMRMHLLR"
 15353..15742
 /note="cognate B of LTR flanking retrotransposon from
 position 10402 to 15747; cognate A resides from position
 10407 to 10796"
 15743..15747
 /note="function-putative-target duplication site; cognate
 B of target duplication site flanking retrotransposon from
 position 10402 to 15747; cognate A resides from position
 10402 to 10406"
 /rpt_type="flanking"
 17025..17085
 /note="function-mitochondrial_repeat; similar to TTV
 element; GenBank accession number X91202"
 /rpt_type="dispersed"
 complement(119590..19593)
 /note="function-putative-target duplication site; cognate
 A of target duplication site flanking retrotransposon from
 position 19590 to 25423; cognate B resides from position
 25420 to 25423"
 /rpt_type="flanking"
 complement(119590..25423)
 /note="function-putative-retrotransposon"
 /rpt_type="dispersed"
 complement(119594..20140)
 /note="cognate A of LTR flanking retrotransposon from
 position 19590 to 25423; cognate B resides from position
 24873 to 25419"
 complement(21779..24805)
 /gene="T26N6.3"
 /note="encodes putative transposon protein; gene model
 last edited on 1 May 99"
 /evidence="not_experimental"
 complement(21779..24805)
 /note="functional catalog ID=90"
 /codon_start=1
 /evidence="not_experimental"
 /product="putative transposon protein"
 /protein_id="AAD29754.1"
 /db_xref="PDB:94773881"
 /db_xref="GI:4773881"
 /translation="MDYKPEFVAGKTIINLEKGVYGHMKVAKRLIRIGLKEAWIAT
 IGKRAPIYKGDGEDYLTQDMQDAEAKRANKRSLSTFNVNQNORFQNCES
 AKEMDKLAKAYEGTSSVKRSRIDMLASQFENLSMEETEEERSGKISATSAEHHNL


```

/codon_start=1
/product="3-isopropylmalate dehydratase"
/protein_id="CAB16402.1"
/db_xref="PID:e349388"
/db_xref="PID:g2414659"
/db_xref="GI:2414659"
/db_xref="SWISS-PROT:014289"
/translation="MSPSYASPKTITKDWMSHYVDLQEDGTCILYIDRHLIHEVTSPOAEGLRTAGRVVRHPELALAVDHNIPDPKDKDIASFHPQDSRTQVATLNNIKVEGLTYGMDNRQGLVHVIGPDEGFTLPGTLYCGSHITSGAFGLARIGTSEVEHVALQTILOKRSKNRIYVNGKLPESIASKDLILHIIIGTAGGGSVIECGEALIESMEARMSCNMSIEAGARAGMIADATTFEYVKNRPPLAPGDDMEQAAVAKTIRSDENAKYDIEVEINADVLPTVMGTSPODVPIPIINGNIPDAVYKDMQAAVAKTLEHMGAKPRTSIVTPIDKVFIGSCTNSIEDRLAAAYKRAKAAVYKDMQAAVAKTGLVKMAEAEGLDQILEAGFDWRBAGSCMGLGNPDOLKPYERCASISNNEEGRQGLAKGRTHLVSPAAAAAIIKSHLCNVEFEFGDVNSGPSIITKNNDPSHVEDGIGLVDDADAVYADAGIATNVAGSVSSAGIPKFTVEGIAAPLPMANVDPDKIIPROFLKTKRRTGLGQFAFYERIRYADGKEIIPDFVLAINEPRHATVLAHNNFGCGSSREHAPVALLNDGCIYIILNPSFADIFENNCFKNGMIPRTPIQVNDMKAAENQKFSYDVLNOITITGDKQYKEVEFEFRKHCVLVNGLDIGLITLQKEMTDAFEAENEPMWNIRSRARLSPKSKNKQSSSRNDW"
/misc_feature
/gene="SPAC9E9.03"
/feature="PS00450 Aconitase family signature"
/feature="8335..9186"
/gene="SPAC9E9.04"
/feature="join(8335..8490,8776..9186)"
/gene="SPAC9E9.04"
/feature="note="SPAC9E9.04", len:188, SIMILARITY: Saccharomyces cerevisiae, 007451, chromosome iv reading frameorf yd1072c., (203 aa), fasta scores: opt: 262, E(): 3.7e-15, (33.7% identity in 140 aa)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB16403.1"
/db_xref="PID:e349389"
/db_xref="PID:g2414660"
/db_xref="GI:2414660"
/db_xref="SPTREMBL:014290"
/translation="MTIYIIVFMILMWEIVSVLLSLPLKVRARILAINASNPFAGRHVKLTICITLILFADSVRYRVKREYDLAIAPSTESARSGYKAAQFYAORNLVIGSALFLFLVYVNRXYLAEMIAADKQALQCTQVEASTNNAKVAEELETLTKLETDRKEVETLEAKYAATKYVEKKDI"
/misc_feature
/gene="SPAC9E9.04"
/feature="note="gtaagg, splice donor sequence"
/feature="8763..8775"
/gene="SPAC9E9.04"
/feature="note="ctaactcttag, splice branch and acceptor"
/feature="9718..10659"
/gene="SPAC9E9.05"
/feature="note="SPAC9E9.05", len:313"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB16404.1"
/db_xref="PID:e349608"
/db_xref="PID:g2414661"
/db_xref="GI:2414661"
/db_xref="SPTREMBL:014291"
/translation="MDSDFVTKAHVITSTPENKLSHRKSVETIYPPSSSNDPFGFSSTKGRILSSINSNDLVNTLSKGTENSNNSYNILPSSPTLEIGIDNEALQIRADENQOQVPTPINSASPSTPLPSSSPLLPKNSSSSPILSLISTSSQQRITPENLNTSKPDSKQLERLIPVHGHLRLKRRRRDDIDLSGLYETKSSPPIHSPEDPSYSDIARSYPKSAFNLRKRGYKREKKIITKYVSQKQDASDNDNTGSGDEENDNKLKELPGKKEYIKSKYFQVDVYQLHVNKG"
/gene="thrc"
/feature="complement(11123..12725)"
/gene="thrc"
/feature="complement(join(11123..12598,12657..12725))"
/gene="thrc"
/feature="note="SPAC9E9.06c, len:514"
/codon_start=1

```

```

/product="threonine synthase"
/protein_id="CAB16415.1"
/db_xref="PID:e1198281"
/db_xref="PID:g2656015"
/db_xref="GI:2656015"
/db_xref="SWISS-PROT:042598"
/translation="MSQVSYLSTGSGSSNSFEPAVYKGLANDGGLPISEPIPLPSGNTIEMKDKSPFEIATFEVMSLTIPRSEISADELKLVDKRSISTFRHPTTLPKSLKNGINLVLEFPGPFEAFEDVALQELGNIJFEFFLRKKNKPEDEDERDLTVGATSGDTGSAAIYGRGRKNDVSFELIPNGRSPLOEAMTVDIPNVHCITVNGVFPDQDLVYKQIFGDVEENKHHIAGVNSIMMARLSLOITLYLSVYKQKADVRFVPPNGFDILAGYVAKRMGLPEKQIVANTNENDILNREFKGRYKADSTOVSPSGPIASKEYSAPAMDIVSNSEBRIYMIATLAEARNHHPAASLSEKWMNEFRKDGVTYVTPVLEAARDIVSESVSNDETIDAIKITYESDHYITIDPHRAGVETGLRCEKTKDDIYITCLSTARHPAKDKVNMALSSYSYDNFTQVLPLEFDGLDEERTCLFSGKPNIDIKOIIIEVTLISREKA"
/misc_feature
/gene="thrc"
/feature="complement(12305..12349)"
/feature="note="PS00165 Serine/threonine dehydratases pyridoxal-phosphate attachment site"
/feature="complement(12599..12619)"
/gene="thrc"
/feature="note="ctaactgcatcttag, splice branch and acceptor"
/feature="complement(12651..12656)"
/gene="thrc"
/feature="note="gtaggt, splice donor sequence"
/feature="complement(14172..14774)"
/gene="yp12"
/feature="complement(14172..14774)"
/gene="yp12"
/feature="note="SPAC9E9.07c, len:200"
/codon_start=1
/product="ypt1-related protein 2."
/protein_id="CAB16405.1"
/db_xref="PID:e349391"
/db_xref="PID:g2414662"
/db_xref="GI:2414662"
/db_xref="SWISS-PROT:P17609"
/translation="MSTKSYDYLIRLLKIGDSGVKSCQLLNFSDSFTPTPTIGIDFKITIEIDGKRIKIQIMDVGDERPRTITVAYRGAMGILLLDVYDVKSKFDVVRTWFSNVQHASSEVYKILGNKDCSDQDQSVSEGOALADEUGVFLFASAKATNVNDFAFPLAREIKKOKIDANEPENQANNNDLGDRIYKRC"
/misc_feature
/feature="complement(14706..14729)"

```

misc_feature

Query Match 60.6%; Score 21.2; DB 7; Length 38078;
 Best Local Similarity 76.5%; Pred. No. 28;
 Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 TTTAAGCTTAGGCATTCTTAAGTCACTATCATCC 34
 Db 10971 TTTAAGCTTAATCTTGAAGACACTTCATCC 10938

RESULT 12

AC005350	LOCUS	AC005350	78661 bp	DNA	PRI	31-JUL-1998
DEFINITION	Homo sapiens chromosome 5, BAC clone 261j17 (LBNT.H190), complete sequence.					
ACCESSION	AC005350					
NID	9336562					
VERSION	AC005350.1					
KEYWORDS	HTG.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.					
AUTHORS	Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K. S., Gunning, K. M., Kadner, K., Miguel, T., Miller, C., Pitluck, S., Pollard, M., Rojeski, H., Subramanian, S. and Martin, C. H.					
TITLE	Sequencing of human chromosome 5					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 78661)					

AUTHORS	REFERENCE	JOURNAL	TITLE	COMMENT	FEATURES
Ricke,D.O.	Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System	Unpublished			source
3 (bases 1 to 78661)					
Kimmerly,M., Bordoc,M., Cheng,J., Connolly,K.S., Gunning,K.M., Davis,C.A., Kaderof,K., Miguel,T., Plittuck,S., Pollard,M., Rojestki,H., Subramanian,S. and Martin,C.H.					
Direct Submission					
Submitted (31-JUL-1998)					
Human Genome Center, DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, MS 74-157, Berkeley, CA 94720, U.S.A.					
Sequence submitted by:					
DOE Joint Genome Institute.					
Location/Qualifiers					
1..78661					
/organism="Homo sapiens"					
/db_xref="taxon:9606"					
/map="5q"					
/clone="261j17"					
/chromosome="5"					
/note="LBNL H190"					
476..656					
/rpt_family="Alu"					
2704..2839					
/note="GRAIL 2 excellent exon, frame 0"					
4504..4592					
/rpt_family="MIR"					
8089..8226					
/rpt_family="Alu"					
complement(8516..9100)					
/rpt_family="MLT2B2"					
9388..10759					
/rpt_family="L1"					
complement(10780..14299)					
/rpt_family="L1"					
complement(15044..15135)					
/rpt_family="MIR"					
complement(16072..16399)					
/rpt_family="LINE"					
19038..19075					
/note="(AT)19"					
/rpt_type=tandem					
/rpt_unit=AT					
complement(19076..19390)					
/rpt_family="Alu"					
23689..24660					
/rpt_family="L1"					
24716..24995					
/rpt_family="Alu"					
25177..27648					
/rpt_family="L1"					
26615..26635					
/note="(T)21"					
/rpt_type=tandem					
/rpt_unit=T					
complement(26618..26908)					
/rpt_family="Alu"					
complement(29199..29244)					
/note="GRAIL 2 excellent exon, frame 2"					
complement(30575..30784)					
/note="GRAIL 2 excellent exon, frame 0"					
31846..31869					
/note="(TA)12"					
/rpt_type=tandem					
/rpt_unit=TA					
complement(34387..34549)					
/rpt_family="MER3"					
35029..35056					
/note="(T)28"					
/rpt_type=tandem					
/rpt_unit=T					
complement(35039..35334)					

```

repeat_region      /rpt_family="Alu"
                   36735..36764
                   /note="(AC)15"
                   /rpt_type= tandem
                   /rpt_unit= AC
misc_feature       complement(39057..39133)
                   /note="GRAIL 2 excellent exon, frame 0"
repeat_region      39883..40023
                   /rpt_family="L1"
                   /rpt_type= tandem
repeat_region      complement(40416..40487)
                   /rpt_family="MER44C"
repeat_region      41142..41292
                   /rpt_family="MIR"
repeat_region      41623..41837
                   /rpt_family="Alu"
repeat_region      45017..45258
                   /rpt_family="L1"
repeat_region      complement(45953..46019)
                   /rpt_family="MIR2"
repeat_region      46124..46204
                   /rpt_family="MIR"
repeat_region      47776..47817
                   /note="(AC)21"
repeat_region      /rpt_type= tandem
                   /rpt_unit= AG
repeat_region      51783..51857
                   /rpt_family="MLT1"
repeat_region      56543..56706
                   /rpt_family="MLT1"
repeat_region      complement(57418..57744)
                   /rpt_family="Alu"
repeat_region      57577..57597
                   /note="(T)21"
repeat_region      /rpt_type= tandem
                   /rpt_unit= T
repeat_region      complement(63432..63490)
                   /rpt_family="MER6"
repeat_region      complement(65010..66136)
                   /rpt_family="L1"
repeat_region      66131..67648
                   /rpt_family="L1"
repeat_region      complement(66461..66710)
                   /rpt_family="Alu"
repeat_region      69340..69369
                   /note="(GT)15"
repeat_region      /rpt_type= tandem
                   /rpt_unit= GT
repeat_region      complement(70474..70725)
                   /rpt_family="Alu"
misc_feature       complement(72044..72130)
                   /note="GRAIL 2 excellent exon, frame 1"
repeat_region      73330..73522
                   /rpt_family="Alu"
repeat_region      complement(74279..75083)
                   /rpt_family="L1"
repeat_region      75262..75289
                   /note="(AC)14"
repeat_region      /rpt_type= tandem
                   /rpt_unit= AC
repeat_region      75291..76794
                   /rpt_family="L1"
repeat_region      77696..77815
                   /rpt_family="MER57A"
BASE COUNT        25567 a 14357 c 14140 g 24597 t
ORIGIN
Query Match        60.6%; Score 21.2; DB 11; Length 78661;
Best Local Similarity 76.5%; Pred.No. 24;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
1 TTTAGCTTAGCATCTTAAGTCAAGTCACTTCATCC 34

```

26 64143 TTAAAGTGACCAATTAATGAGATCATCC 64176

```

RESULT 13
LOCUS AC006222
DEFINITION Homo sapiens, clone hRPK.12_A.1, complete sequence.
ACCESSION AC006222
MID 94159874
VERSION AC006222.1 GI:4159874
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165643)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone hRPK.12_A.1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165643)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P.,
FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S.,
Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C.,
Jones,C., Kann,L., Karatas,A., Lehoczy,J., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrum,J.,
Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Naylor,J.,
Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,R., Peterson,K.,
Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesiaye,S., Tortunella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL * Submitted (13-DEC-1998) Whitehead Institute/MIT Center for Genome
REFERENCE Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS 3 (bases 1 to 165643)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckery,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Lieu,C., Locke,K., MacDonald,P.,
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McDonald,P.,
Meldrum,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,
Tesiaye,S., Tortunella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (15-JAN-1999) Whitehead Institute/MIT Center for Genome
COMMENT Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 15, 1999 this sequence version replaced gi:4153849.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html.
FEATURES
source location/Qualifiers
1..165643
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="hRPK.12_A.1"
/clone_lib="RPCI-11 human BAC library"
repeat_region 111..2188
/rpt_family="L1PA12"
repeat_region 2185..2305
/rpt_family="L1M2"
repeat_region 2320..2427
/rpt_family="L1P"
repeat_region 2433..5941

```

```

repeat_region complement(6541..6645)
/rpt_family="L1PA13"
repeat_region complement(6647..6753)
/rpt_family="MSTD"
repeat_region complement(6647..6753)
/rpt_family="MSTD"
repeat_region complement(7652..7836)
/rpt_family="L2"
repeat_region complement(8505..8669)
/rpt_family="MLT1J1"
repeat_region 8717..8743
/rpt_family="TTTA)n"
repeat_region complement(8747..9035)
/rpt_family="L1M3A"
repeat_region complement(9048..9477)
/rpt_family="L1PA7"
repeat_region complement(9524..9585)
/rpt_family="MLT1J1"
repeat_region 9956..9991
/rpt_family="AT-rich"
repeat_region 10432..10716
/rpt_family="L1M3A"
repeat_region 10968..11016
/rpt_family="AT-rich"
repeat_region 11120..11152
/rpt_family="AT-rich"
repeat_region 11595..11970
/rpt_family="L1M3A"
repeat_region 12166..12264
/rpt_family="MLT1H"
repeat_region 12419..12687
/rpt_family="ALUSg"
repeat_region complement(12898..13042)
/rpt_family="MIR"
repeat_region 13417..13508
/rpt_family="GA-rich"
repeat_region 13491..13849
/rpt_family="(GGA)A)n"
repeat_region 13598..13653
/note="Single-stranded terminator coverage."
repeat_region 13839..13906
/rpt_family="GA-rich"
repeat_region complement(14648..14707)
/rpt_family="L2"
repeat_region 16861..17001
/rpt_family="MLT1A1"
repeat_region 17385..17439
/rpt_family="MLT1H"
repeat_region complement(17652..18126)
/rpt_family="MLT1D"
repeat_region 18170..18231
/rpt_family="L1M3A"
repeat_region complement(21556..21881)
/rpt_family="L1M3A"
repeat_region 22385..22592
/rpt_family="MIR"
repeat_region 22603..22668
/rpt_family="MER81"
repeat_region 22901..23033
/rpt_family="MER5B"
repeat_region 23204..23232
/rpt_family="(TTT)G)n"
repeat_region 24016..24045
/rpt_family="(TTTA)n"
repeat_region complement(24046..24373)
/rpt_family="L1PA9"
repeat_region 24621..24707
/rpt_family="AT-rich"
repeat_region 24764..24795
/rpt_family="(TTTA)n"
repeat_region complement(24796..25079)
/rpt_family="ALUSp"
repeat_region 26220..26243
/rpt_family="AT-rich"

```

FEATURES	source
ACCESSION	U09810
NID	g557273
VERSION	U09810.1
KEYWORDS	GI:557273
SOURCE	
ORGANISM	black chiton. Mitochondrion: Katharina tunicata Eukaryota; Metazoa; Mollusca; Polyplacophora; Neoloricata; Ischnochitonidae; Mopallidae; Katharina.
REFERENCE	1 (bases 1 to 15332) Boore, J.L. and Brown, W.M. Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina tunicata
AUTHORS	Genetics 138 (2), 423-443 (1994)
TITLE	95129806 2 (bases 1 to 15332) Boore, J.L.
JOURNAL	Direct Submission Submitted (17-MAY-1994) Jeffrey L. Boore, Cell Biology and Neuroanatomy, University of Minnesota, 321 Church St. SE, Minneapolis, MN 55455, USA
REFERENCE	Location/Qualifiers 1. .15332 /organism="Katharina tunicata" /mitochondrion /db_xref="taxon:34587" 1. .1548 /note="followed by 'T' as an abbreviated stop codon" /codon_start=1 /transl_table=5 /codon=(seq:"tga", aa:Trp) /product="cytochrome oxidase subunit I" /protein_id="AAC48362.1" /db_xref="PID:g557274" /db_xref="GI:557274" /translation="MFMFSTNHRKDIGTLYILFGIAGLVGTALSLIRAELEGOPAL LGSDOLNIVTAAHAFMIFELVPMKIGFGMLVPLDMGPDMAFPRLNNSFWLL PPALCILLAGAIVESGAGCTGVTYPRAGVAGAGSYDAIFSLHAGYSILGAVN FITTYNMRSEGMQLERLPFPWSVKITALLISLPLVLAGGITMLTDNPNFTSPD PAGGDDPILYLQHLPEFGHREYVILILPGFGLSHLVHYSKKEFTFGLGMVYAMA IGLIGLIVAHNMFVVGMDVIRAYFAFTAAIMTIAVPGLIFPEMLATIVGARIGSERP MMALGIFELFTEVGLGLTGLVINSISLIDIMLHDSYVVAHRYVLSMGAVFALGAEY WYPLDGLSLSEHRKTSHEFVVMFLGVNLLPEFGHFLGSGMPRYSPPCYIKWNV SSGISLSPFAVLEFMEIWESELVSQRGVYWSHLSALEMNSIPUGFTDLMKGLL FVNLLIL"
CDS	1519..1682 /note="codon recognized: GAC" /product="trna-asp" /anticodon=(pos:1650..1652,aa:Asp) 2107..2796 /codon_start=1 /transl_table=5 /codon=(seq:"tga", aa:Trp) /product="cytochrome oxidase subunit II" /protein_id="AAC48365.1" /db_xref="PID:g557277" /db_xref="GI:557277" /translation="MAFWSQWGFQDASPIEQLIEFHDAMLILMIISLSYGAVS LNNPSRSTLESQELIETWILIPAVVILFLAFPSQLLYLDELSEPALITKVVG QWTVSEYSDLELFDPSVLSLEDEEGDYRLLEVDHRSVPMKTRVRLVYAADVL HSTVLSIGYKADAVPGRNLQSLSFANYPOVFTGQCSLEIGANHSMPITYLEVYDSS FIRWIMFNGEA" 2872..3033 /codon_start=1 /transl_except=(pos:3001..3003,aa:Trp) /transl_table=5 /product="ATPase subunit 8" /protein_id="AAC48364.1" /db_xref="PID:g3335323" /db_xref="GI:3335323" /translation="MPQLAPMNMWIFLLTFWNSVFCLVGVMMVKSSSYSLSKSVSM PFYFKRWVM" 3037..3729 /codon_start=1

/transl_table=5
/product="ATPase subunit 6"
/protein_id="AAC48363.1"
/db_xref="PID:g557275"
/db_xref="GI:557275"
/translation="MADDFSSFDNNFTLSFYLIMVLGLAVLISQNMFKSKR
FYSLPFPKNMGQVWARKNISKSGFINLWPSGGLINPLVIVETYSISVRPTLSY
FARFCGLPFLSLFSGDLVYSEVSHLIPSGAGILNPLVIVETYSISVRPTLSY
RLANMSAGHIIIGLIGSYLVSGLCPIVLGLLLEVOVFYFLFELIGVSLIQAFIRS
LVLSYSDHRY"
complement(3758..3822)
/note="codon recognized: TTC"
/product="tRNA-Phe"
/anticodon="(pos:3789..3791,aa:Phe)
complement(3821..3536)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAC48366.1"
/db_xref="PID:g557278"
/db_xref="GI:557278"
/translation="MKMPLSTQATVILLSHSLSESTAMVTLTKIILLECFFN
FNSYITLPIINVTALFSAIVCYISAVMIFPKNMOSODIFIKRFTWIMLVLSM
NLIFPSIALTLGMDGIGVSRCLVYQNTKSLAGMTAFMRNLGDVAILLATS
VTSGQHFNALFMSSTPTMTITISMIILAGTRSAQLPFCSMPLAAMAAPPVVALVH
SSVLVAGVELLIQFVPLSNHLSFTSLMITSMTVMAGISANFETDKITIALST
LSQGVNMLSVSLGFLSLALFLTHAMFKALFLPCAGNIHSHNNODIKMHWMT
QMPSTCFNANLALGCFPMAGYKSDVIEEMNQNNFLISLNNFLITLASY
SARSLFPMKSSSVSTSDSDNSIAPLTLAGAITSAGVFSMLILTPSTLPL
ISMNRKMTAPATLTGAATYKLEKLEKLSHSTINFFEMEMINISNMTRKSSSL
YTGITFKSLDGMETIGGBGLNKTMTTITINOKNOSNLNPLFISLTLISMLPISV
TMTT"
complement(5537..5601)
/note="codon recognized: CAC"
/product="tRNA-His"
/anticodon="(pos:5568..5570,aa:His)
complement(5597..6928)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_except="(pos:complement(6788..6790),aa:Trp)
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAC48367.1"
/db_xref="PID:g557280"
/db_xref="GI:557280"
/translation="MSAFSTVMTKSKMYKMSISHEFFMTLNEFSPMNPKIN
YEMSDNEFVPLMLSCWITGLMINSYKILKNNLVSEFLNMTINFTIIMVETOK
SLPSLYEFESALIPLLIILMGYOPEROAGMYMTITIGALPLINFTIYKSHN
AHLNLLMMSLPIPIYQAMISFHWLITILVLYLPIYFHLMPKHAVEAPVAGSMI
LAULLKLGIGINTINNIRKYNVLYFTITGLVGVVSLICLRQSPKALAY
SSVGHMGLMAGLISFEFGKMLMLMAGLSSGIFSNMNMKYESGGRSFLFK
GFLSVPTFSMFLMCSINNAAPSIINLSEAGMISTVSIKFMILMLMVFSA
VYTLFTNTHGHSISFVNPNENKPLVYVIFLHMIPALLILLSTTC"
complement(6940..7242)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4L"
/protein_id="AAC48368.1"
/db_xref="PID:g557280"
/db_xref="GI:557280"
/translation="MANFNPFLTGLALSSISALMTICLQKHLNLTLLALEITLMSI
FLLVLSLSNLSNGLMIFVLLTLAASEASIGLMLVILLIRAHNDVYSSISIRKC"
7247..7312
/note="codon recognized: ACA"
/product="tRNA-Thr"
/anticodon="(pos:7275..7277,aa:Thr)
complement(7312..7378)
/note="codon recognized: TCA"
/product="tRNA-Ser"
/anticodon="(pos:7348..7350,aa:Ser)
complement(7381..8520)
/codon_start=1

/transl_except="(pos:complement(8176..8178),aa:Trp)
/transl_table=5
/product="cytochrome b"
/protein_id="AAC48371.1"
/db_xref="PID:g557281"
/db_xref="GI:557281"
/translation="MKPIRSQVPEKINGSLIDPSPNSLITMWNFSLGLCLIT
OITGLFLAHYATNITELASSVAHICRDVNYGCLTAAHANGSLFFICLFRIGG
MYGSEFNITWVWVILFLPMTGTAIGVLAAPVGMGSPWGAIVITLMSAIPYIGTT
MYQWINGCFVADNATLTFPSEFHLAPFIIAASVHLILFLETGSSNPLGINDSER
VTIHSEFTKDLGVVAVIFLLVFLPDLNLAHPNFIAPNLTVPPTHPPEMYEL
FAVALISPNKIGGVGLMSIVILEFVPLINFGKRTISFVPTQILVMSLVSVEL
LTMWAGQVPEDEPILGQITTFYFLFLLPIPTKMMWMTINE"
complement(8513..9019)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/protein_id="AAC48369.1"
/db_xref="PID:g557281"
/db_xref="GI:557281"
/translation="MFLPLMSLIFSCAFSLPLTQPIISGLLPLNSLPLISILPFG
SSTWSEFLFLIYIGLIVMAYVATLAMPNLMFKRNLQVFFPVSTVFMILMLTSDP
IGTLEROTKMFIMNSFPHLIGLSLSPNLLIIVSLALILFVLICVYKICVFSNG
PLRPYKXA"
9024..9089
/note="codon recognized: CCA"
/product="tRNA-Pro"
/anticodon="(pos:9054..9056,aa:Pro)
complement(9082..10032)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC48370.1"
/db_xref="PID:g557282"
/db_xref="GI:557282"
/translation="MFLSPSMNVAYISILLAVFLLERKGLGYTORGRPKVSI
FGLAQPLADKLFLKEENNYNSNKLPIYAPVLSILSLCLSLVSDPTLWTFNF
GVLEFICTSALSVYGTMIAGWASNSKALGLAEVAVQTSYETLVLVLSAALS
SLNINQISHSQHLWNSILMTPLIMMIVVCLAEVNAPEDFAGESELVSGVNIERG
SGSFALLFLAEVASILFMSMNTTILFAGAGSWILTNFLVYKVTFAFVIVRGTF
PLRVDLIMKLTWISFLPVSIMTILMMLMSAKPMVTILLSS"
complement(10033..10101)
/note="codon recognized: TTA"
/product="tRNA-Leu"
/anticodon="(pos:10068..10070,aa:Leu)
complement(10101..10165)
/note="codon recognized: CTA"
/product="tRNA-Leu"
/anticodon="(pos:10132..10134,aa:Leu)
complement(10166..11440)
/note="The exact endpoints are uncertain"

Query Match 60.6% Score 21.2: DB 37; Length 15532;
Best local similarity 76.5% Pred. No. 34;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 1 TTTAGCTTAGGCATTCTAAGTCACTATCATCC 34
||||||| ||| ||||| ||| |||||
DB 10306 TTTAGCTTCGACATTAAAGTTACTTATCC 10339

RESULT 15
SLH4K SLH4K 1673 bp DNA INV 12-SEP-1993
LOCUS STYLOMYCHIA lemae histone H4 gene (clone H4K).
DEFINITION X16018
ACCESSION X16018
NID 910141
VERSION X16018.1 GI:10141
KEYWORDS histone; histone H4.
SOURCE Stylomychia lemae.
ORGANISM Stylomychia lemae
Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichidae;

```

REFERENCE      Oxytrichidae; Stylonychia.
AUTHORS        1 (bases 1 to 1673)
TITLE          Direct Submission
JOURNAL        Submitted (03-AUG-1989) Wefes I.,
               Medizinische-Naturwissenschaftliches Forschungszentrum,
               Universitaet Tuebingen, Ob dem Himmelreich 7, 7400 Tuebingen, F R G
REFERENCE      2 (bases 1 to 1673)
AUTHORS        Wefes, I. and Lipps, H.J.
TITLE          The two macronuclear histone H4 genes of the hypotrichous ciliate
               Stylonychia lemnae
JOURNAL        DNA Seq. 1 (1), 25-32 (1990)
MEDLINE        92119316
COMMENT        Data kindly reviewed (19-NOV-1990) by Lipps H.J.
FEATURES       Location/Qualifiers
               source
               1..1673
               /organism="Stylonychia lemnae"
               /strain="EK16Sp"
               /db_xref="taxon:5949"
               /clone="H4K"
               1..20
               /note="telomeric sequence"
               misc_feature 365..401
               /note="ARS sequence"
               misc_feature 785..821
               /note="ARS sequence"
               misc_feature complement(1014..1024)
               /note="ARS sequence"
               misc_feature 1072..1103
               /note="ARS sequence"
               promoter 1117..1121
               /note="CAAT-box"
               promoter 1189..1195
               /note="TATA-box"
               CDS 1208..1522
               /note="histone H4"
               /codon_start=1
               /transl_table=6
               /protein_id="CA34151.1"
               /db_xref="PID:910142"
               /db_xref="GI:10142"
               /translation="MAGRGKVGKGYGAKRHTKSLKETIMGITKPAIRRLARRG
               VKRISSLIYEERNYLRSLFENVIRDSVYTFEHAARKTYTALDVIYALKRQRTLXGF
               GG"
               misc_feature 1557..1588
               /note="ARS sequence"
               misc_feature 1576..1598
               /note="ARS sequence"
               misc_feature 1654..1673
               /note="telomeric sequence"
BASE COUNT    606 a 276 c 266 g 525 t
ORIGIN
Query Match      59.4%; Score 20.8; DB 36; Length 1673;
Best Local Similarity 78.1%; Pred. No. 79;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY      1 TTAAAGCTTAGCATCTTAAGTCACTATCAT 32
        |||||  |||||  |||||  |||||  ||
DB      1163 TTAAAGCATGTCATTTAAAGTCTCTATAT 1194

```

Search completed: September 13, 1999, 15:55:23
 Job time: 4551 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:30 ; Search time 255.05 Seconds

(without alignments)
34.333 Million cell updates/sec

Title: US-09-325-095-17

Perfect score: 35
Sequence: 1 TTTAGCTTAGGCATTCTTAAGTCATCATCC 35

Scoring table: IDENTITY_MDC

Searched: 31585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	35	1	059016
2	35	100.0	35	1	057022
3	26	74.3	3207	1	051155
4	26	74.3	3498	1	057012
5	24.4	69.7	3412	1	051156
6	19.8	56.6	314	1	077252
7	19.8	56.6	10669	1	052190
8	19.6	56.0	1196	1	T62860
9	19.6	56.0	1196	1	T62861
10	19.2	54.9	400	1	V62749
11	19.2	54.9	400	1	V78144
12	19.2	54.9	344	1	V78616
13	19.2	54.9	13856	1	V74342
14	19.2	54.9	2821	1	X13998
15	18.8	53.7	240	1	T24992
16	18.6	53.1	11531	1	054222
17	18.6	53.1	1378	1	045241
18	18.6	53.1	1474	1	065403
19	18.6	53.1	1474	1	092304
20	18.6	53.1	1515	1	T07072
21	18.6	53.1	1474	1	V15238
22	18.6	53.1	1320	1	X05214
23	18.6	53.1	1320	1	X05215
24	18.6	53.1	110000	1	X20248_06
25	18.6	53.1	111309	1	X20250
26	18.6	53.1	4516	1	X26856
27	18.6	53.1	1360	1	X33954
28	18.4	52.6	3450	1	T85113
29	18.4	52.6	9058	1	V10663
30	18.2	52.0	18177	1	N50490
31	18.2	52.0	1059	1	094068
32	18.2	52.0	564	1	V74801
33	18.2	52.0	18303	1	X04502
34	18.2	51.4	243	1	T03852
35	18.2	51.4	8878	1	V52380
36	18.2	51.4	1805	1	V21209_03
37	18.2	51.4	110000	1	V21209_16
38	17.8	50.9	64976	1	X13283
39	17.8	50.9	2966	1	X13283
40	17.6	50.3	4456	1	092882
41	17.6	50.3	4841	1	T05696
42	17.6	50.3	459	1	T61415
43	17.6	50.3	3754	1	V74865

ALIGNMENTS

RESULT 1	059016	059016 standard; DNA; 35 BP.	
AC	059016:	12-APR-1994 (first entry)	
DE	RACE PCR primer 2280.		
KW	Phosphoinositide kinase; PI; p85 subunit; screening; agonist;		
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;		
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;		
KW	ss; amplification; p110.		
OS	Synthetic.		
PN	W0321328-A.		
PD	28-OCT-1993.		
PF	13-APR-1993; G00761.		
PR	13-APR-1992; GB-008135.		
PA	(LUDW.) LUDWIG INST CANCER RES.		
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;		
PI	Patkerp, Volinia S, Waterfield MD;		
DR	WPI; 93-351738/44.		
PR	activity, useful for controlling cell proliferation		
PS	Recombinant polypeptide(s) - with phosphoinositide-3 kinase		
CC	Ad SGBAF-1 cell line was established by transfection of bovine		
CC	adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was		
CC	isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR		
CC	was performed by synthesizing first strand cDNA from random hexamers		
CC	on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-		
CC	dr as primers. Products were fractionated using an agarose gel. The		
CC	DNA was isolated from the gel and subjected to PCR using oligo 2280 and		
CC	adaptor as primers. A product of 350 bp was further sequenced.		
CC	See also 051155-6, 059012-23 and 057522-3.		
SO	Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;		
Query Match	100.0%; Score 35; DB 1; Length 35;		
Best Local Similarity	100.0%; Pred. NO. 2e-06;		
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Oy	1 TTTAGCTTAGGCATTCTTAAGTCATCATCC 35		
DB	1 TTTAGCTTAGGCATTCTTAAGTCATCATCC 35		
RESULT 2	057022	057022 standard; DNA; 35 BP.	
AC	057022:	31-AUG-1994 (first entry)	
DE	Ptdins 3-kinase oligo 2280 primer.		
KW	110 kD catalytic subunit; phosphatidylinositol 3-kinase;		
KW	transformation; schizosaccharomyces pombe; nmt promoter; thiamine;		
KW	Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;		
KW	blood vessel plaques; ss.		
OS	Synthetic.		
PN	W09403609-A.		
PD	17-FEB-1994.		
PF	05-AUG-1993; G01651.		
PR	05-AUG-1992; GB-016654.		
PA	(IMCR.) IMPERIAL CANCER RES TECHNOLOGY.		
PI	Goode NT, Nurse PM, Parker PJ, Waterfield MD;		
DR	WPI; 94-065697/08.		
PR	Eukaryotic cells transformed with mammalian phospholipid or		
PI	protein kinase DNA - useful in assays for compounds involved in		
PI	cell growth regulation and for treating cancers		
PS	Example 1; Page 15; 71pp; English.		
CC	The sequences given in 057020-23 are primers which were used in the		

CC Isolation of the phosphatidyl inositol (PtdIns) 3-kinase cDNA. The
CC amplified sequence was placed under the regulatory control of the nmt
CC promoter. These sequences were transformed into Schizosaccharomyces
CC pombe cells in an embodiment of the invention. In the presence of
CC thiamine the promoter is inactive and the cells carrying the PtdIns
CC plasmids grow as the parental strain. In the absence of thiamine the
CC nmt promoter functions and the PtdIns is induced. PtdIns activity is
CC substantially increased under these conditions. Cells containing
CC constructs such as this, are useful in assays for detecting compounds
CC involved in cell growth regulation. It is also used as the basis for
CC detecting compounds for treating cancers and the formation of blood
CC vessel plaques.

Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;

Query Match 100.0%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 2e-06;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 1 TTTAGCTTAGGCACTTCTAAGTCACTATCATCC 35
|||||
DB 1 TTTAGCTTAGGCACTTCTAAGTCACTATCATCC 35

RESULT 3
Q51155/c
ID Q51155 standard; cDNA; 3207 BP.

AC Q51155;
DT 12-APR-1994 (first entry)

DE p110 cDNA.

KM Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KM platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW 5s.

Key Location/Qualifiers

FT cds 1..3207
FT /tag= a
FT /note= "PI3- kinase p110"

PD WO9321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1993; GB-008135.

PA (LUDWIG-) LUDWIG INST CANCER RES.

PI Dhond R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parker PJ, Volinia S, Waterfield MD;

DR P-PSDB; R43341.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Claim 7; fig 9; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe conty.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51156, Q59012-23 and Q57522-3.

CC Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

DB 110 AGGCATTCTAAGTCACTATCATCC 85

RESULT 4

ID Q57012/c

AC Q57012 standard; cDNA to mRNA; 3498 BP.

DT 31-AUG-1994 (first entry)

DE PtdIns 3-kinase 110 KD catalytic subunit cDNA.

KM 110 KD catalytic subunit; phosphatidyl inositol 3-kinase;

KM transfection; Schizosaccharomyces pombe; nmt promoter; thiamine;

KM PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;

KW blood vessel plaques; 5s.

OS Bos taurus.

Key Location/Qualifiers

FT cds 1..3207
FT /tag= a
FT /product= p110

PN WO9403609-A.

PD 17-FEB-1994.

PF 05-AUG-1993; G01651.

PR 05-AUG-1993; GB-016654.

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;

DR WPI; 94-065697/08.

DR P-PSDB; R46294.

PT Eukaryotic cells transformed with mammalian phospholipid or

PT protein kinase DNA - useful in assays for compounds involved in

PT cell growth regulation and for treating cancers

PS Disclosure: fig 1; 71pp; English.

CC This sequence encodes the 110 KD catalytic subunit of the phosphatidyl

CC inositol (PtdIns) 3-kinase. This sequence was transformed into

CC Schizosaccharomyces pombe cells under the regulatory control of the

CC nmt promoter in an embodiment of the invention. In the presence of

CC thiamine the promoter is inactive and the cells carrying the PtdIns

CC catalytic subunit plasmid grow as the parental strain. In the absence

CC of thiamine the nmt promoter functions and the PtdIns 3-kinase

CC catalytic subunit is induced. PtdIns activity is substantially

CC increased under these conditions. Cells containing constructs such as

CC this, are useful in assays for detecting compounds involved in cell

CC growth regulation. It is also used as the basis for detecting

CC compounds for treating cancers and the formation of blood vessel

CC plaques.

CC Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 74.3%; Score 26; DB 1; Length 3498;

Best Local Similarity 100.0%; Pred. No. 0.024;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 10 AGGCATTCTAAGTCACTATCATCC 35
|||||
DB 110 AGGCATTCTAAGTCACTATCATCC 85

RESULT 5

ID Q51156/c

AC Q51156 standard; cDNA; 3412 BP.

DT 12-APR-1994 (first entry)

DE Human p110 cDNA.

KM Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ds.

OS Human.

Key Location/Qualifiers

FT cds 1..3207
FT /tag= a
FT /note= "PI3- kinase p110"

PN WO9321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

Query Match 74.3%; Score 26; DB 1; Length 3207;

Best Local Similarity 100.0%; Pred. No. 0.023;

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q7 10 AGGCATTCTAAGTCACTATCATCC 35
|||||

PR 13-APR-1992: GB-008135.
PA (LUDW.) LUDWIG INST CANCER RES.
PI Dhand R, Fry MO, Gout I, Hiles ID, Otsu M, Panayotou G;
PI ParkerPJ, Volinia S, Waterfield MD;
DR WPI: 93-351738/44.
P-BSDB: R43342.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity, useful for controlling cell proliferation
PS Claim 7; Fig 16; 146pp; English.
CC Southern blot analysis was performed using a bovine cDNA probe contg.
CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated
CC from a cDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC p13 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with p13 kinase activity and is useful for
CC screening for (anti)agonists of p13 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also Q51155 and Q57522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 69.7%; Score 24.4; DB 1; Length 3412;
Best Local Similarity 96.2%; Pred. No. 0.11;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 10 AGGCATTCTAAGTCATCATCATCC 35
DB 110 AGGCATTCTAAGTCATCATCATCC 85

RESULT 6
077252/C
ID 077252 standard; DNA: 314 BP.
AC 077252;
DT 23-SEP-1994 (first entry)
DE Human genome fragment (Preferred).
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
PS Homo sapiens.
PN W09401548-A.
PD 20-JAN-1994.
PF 13-JUL-1993: G01467.
PR 13-JUL-1992: GB-014857.
RA (MEDT.) MEDICAL RES COUNCIL.
PI Gross J, Hadfield KM, Howells D, Kelly M, Shaw D.
PI Silson DR, Starkey M;
DR WPI: 94-035056/04.
PT New nucleic acid fragment encoding gene products - can be used
PT for genetic analysis and mapping
CC Claim 1; Page 429; 616pp; English.
CC human nucleic acid fragments, isolated from brain, adrenal tissue,
CC the placenta or bone marrow comprise any of: (A) a sequence
CC selected from (076401-077613), (B) an allelic variation of a
CC sequence as described in (A), or (C) a sequence complementary
CC to (A) or (B).
CC Preferred sequences exhibit no more than 90% homology to a human
CC sequence known per se.
SV Sequence 314 BP; 96 A; 65 C; 58 G; 95 T;

Query Match 56.6%; Score 19.8; DB 1; Length 314;
Best Local Similarity 77.4%; Pred. No. 7.3;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 TTAAGCTTAGGATTCTAAGTCATCATCAT 32
DB 138 TTAAGCTTAGGATTCTCAAAACACTTTCAT 108

RESULT 7
ID V52190 standard; DNA: 10669 BP.
AC V52190;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:57.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN W09818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-272225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae.
PS Claim 1; Page 504-510; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridise to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 10669 BP; 3245 A; 1967 C; 2353 G; 3104 T;

Query Match 56.6%; Score 19.8; DB 1; Length 10669;
Best Local Similarity 77.4%; Pred. No. 12;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 4 AAGCTTAGGATTCTAAGTCATCATCATCC 34
DB 2033 AAGCTTAGGATTCTAAGTCATCATCATCC 2063

RESULT 8
ID T62680 standard; DNA: 1196 BP.
AC T62680;
DT 28-MAY-1997 (first entry)
DE Baboon SP-10 CDNA encoding acrosomal sperm antigen 10.
KW SP-10; spermatogenesis; acrosomal sperm antigen; membrane;
KW contraceptive vaccine; testis-specific; differentiation; ss.
OS Papio papio.
FT Key Location/Qualifiers
FT misc_signal 65..71
FT /*tag= a
FT /*note= "5' consensus sequence flanking ATG start"
FT cds 72..929
FT /*tag= b
FT signal_peptide 72..122

```

FT      mat_peptide      /tag- c
FT      123. .926
FT      /tag- d
FT      misc_feature      583. .684
FT      /tag- e
FT      /note- "alternately spliced out"
FT      misc_signal      1000. .1004
FT      /tag- f
FT      /note- "mRNA degradation consensus sequence"
FT      poly_a_signal      1166. .1171
FT      /tag- g
PN      US5602005-A.
PD      11-FEB-1997.
PF      03-MAR-1989. 318551.
PR      03-MAR-1989; US-318551.
PR      16-FEB-1990; US-481491.
PR      27-MAR-1992; US-858798.
PR      18-AUG-1994; US-292045.
PA      (UYYI-) UNIV VIRGINIA PATENTS FOUND.
PI      Herr JC, Wright RM;
DR      MPI; 97-131798/12.
DR      P-PSDB; W14465.
PT      Intra-acrosomal primate sperm antigen - for contraceptive vaccine
PS      prodn.
CC      Example 14; Fig 16; 70pp; English.
CC      The present sequence is SP-10 cDNA isolated from a baboon testis cDNA
CC      library. Alternative splicing is exhibited within the open reading frame,
CC      with a 102 bp internal deletion. The flanking sequence of this deletion
CC      encodes the 5' GTG-GAG 3' consensus splice sequence characteristic of an
CC      intron. The SP-10 protein is an intra-acrosomal primate sperm antigen
CC      that remains associated with the primate sperm after the acrosome
CC      reaction and is bound by monoclonal antibody produced by cell line ATCC
CC      HB10039. The antigen can be used for the production of contraceptive
CC      vaccines.
SQ      Sequence 1196 BP; 307 A; 287 C; 259 G; 343 T;

Query Match      56.0%; Score 19.6; DB 1; Length 1196;
Best Local Similarity 73.5%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      1 TTTAGCTTAGCATCTCTAAGTCACTATCATCC 34
      ||| || ||||| | ||||| ||||| |
DB      949 TTTGACTCAGCAGTAAAGTCTCTATCATTC 982

RESULT 9
ID      T62681 standard; DNA; 1196 BP.
AC      T62681;
DT      28-MAY-1997 (first entry)
DE      Macaque SP-10 cDNA encoding acrosomal sperm antigen 10.
KW      SP-10; spermatogenesis; acrosomal sperm antigen; membrane;
KW      contraceptive vaccine; testis-specific; differentiation; ss.
OS      Macaca fascicularis.
PI      Key
PI      Location/Qualifiers
FT      misc_signal      4
FT      /tag- a
FT      /note- "major transcriptional start site"
FT      65. .71
FT      /tag- b
FT      /note- "5' consensus sequence flanking ATG start"
FT      72. .929
FT      /tag- c
FT      /tag- c
FT      signal_peptide      72. .122
FT      /tag- d
FT      mat_peptide      123. .926
FT      /tag- e
FT      misc_feature      583. .684
FT      /tag- f
FT      /note- "alternately spliced out"
FT      1000. .1004
FT      /tag- g
FT      misc_signal

```

```

FT      poly_a_signal      /note- "mRNA degradation consensus sequence"
FT      1166. .1171
FT      /tag- h
PN      US5602005-A.
PD      11-FEB-1997.
PF      03-MAR-1989. 318551.
PR      03-MAR-1989; US-318551.
PR      16-FEB-1990; US-481491.
PR      27-MAR-1992; US-858798.
PR      18-AUG-1994; US-292045.
PA      (UYYI-) UNIV VIRGINIA PATENTS FOUND.
PI      Herr JC, Wright RM;
DR      MPI; 97-131798/12.
DR      P-PSDB; W14466.
PT      Intra-acrosomal primate sperm antigen - for contraceptive vaccine
PS      prodn.
CC      Example 14; Fig 16; 70pp; English.
CC      The present sequence is SP-10 cDNA isolated from a macaque testis cDNA
CC      library. Alternative splicing is exhibited within the open reading frame,
CC      with a 102 bp internal deletion. The flanking sequence of this deletion
CC      encodes the 5' GTG-GAG 3' consensus splice sequence characteristic of an
CC      intron. The SP-10 protein is an intra-acrosomal primate sperm antigen
CC      that remains associated with the primate sperm after the acrosome
CC      reaction and is bound by monoclonal antibody produced by cell line ATCC
CC      HB10039. The antigen can be used for the production of contraceptive
CC      vaccines.
SQ      Sequence 1196 BP; 304 A; 283 C; 261 G; 348 T;

Query Match      56.0%; Score 19.6; DB 1; Length 1196;
Best Local Similarity 73.5%; Pred. No. 11;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY      1 TTTAGCTTAGCATCTCTAAGTCACTATCATCC 34
      ||| || ||||| | ||||| ||||| |
DB      949 TTTGACTCAGCAGTAAAGTCTCTATCATTC 982

RESULT 10
ID      V62749/c
ID      V62749;
AC      V62749;
DT      15-FEB-1999 (first entry)
DE      Human secreted protein clone er418_5 cDNA.
KW      Secreted protein; human; er418_5; ds.
OS      Homo sapiens.
PI      Key
PI      Location/Qualifiers
FT      CDS      571. .3306
FT      /tag- a
PN      M09846757-A2.
PD      22-OCT-1998.
PF      14-APR-1998; U07999.
PR      13-APR-1998; US-059487.
PR      15-APR-1997; US-843374.
PA      (GEMV) GENETICS INST INC.
PI      Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI      Racine LA, Spaulding V, Treacy M;
DR      MPI; 98-568731/48.
DR      P-PSDB; W74723.
PT      New polynucleotide(s) encoding secreted human proteins - are derived
PT      from, e.g. human foetal brain or foetal kidney cDNA libraries,
PT      potentially useful as, e.g. vaccines or thrombolytic agents
PS      Claim 19; Page 76-78; 120pp; English.
CC      Full-length cDNA clone er418_5 includes an open reading frame
CC      encoding a human secreted protein (see W74723). It was isolated
CC      from a human foetal brain cDNA library using methods which are
CC      selective for cDNAs encoding secreted proteins, or was identified
CC      as encoding a secreted or transmembrane protein on the basis of
CC      computer analysis of the amino acid sequence of the encoded protein.
CC      It shows some similarity to database sequences. The invention
CC      provides polynucleotides (see V62746-55) from human foetal brain,
CC      adult testis, adult brain, adult kidney and foetal kidney (all
CC      deposited as composite clone ATCC 98404), which encode human

```

CC secreted proteins (see W4720-29). The polynucleotides and
 CC proteins are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data are
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating (e.g. as vaccines) or suppressing activity.
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC antiinflammatory activity, cadherin/tumour invasion suppressor
 CC activity, and tumour inhibition activity. The polynucleotides are
 CC also stated to be useful for gene therapy, and can be used in
 CC recombinant production of the polypeptides.
 SQ Sequence 4130 BP; 1113 A; 1051 C; 1024 G; 936 T;

Query Match 56.0%; Score 19.6; DB 1; Length 4130;
 Best Local Similarity 73.5%; Pred. No. 12;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 2 TTAAGCTTAGCATTCTAAAGTCATCATCC 35
 DB 794 TTAATCTTAGCATTTCAAGGACATCCCTTCC 761

RESULT 11

V78144
 ID V78144 standard; DNA; 400 BP.
 AC V78144;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #3833.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 CS Staphylococcus aureus.
 PN EP-786519-N2.
 PD 30-JUL-1997.
 PR 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DT WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1; Page 2708; 3271bp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 400 BP; 114 A; 77 C; 67 G; 139 T;

Query Match 54.9%; Score 19.2; DB 1; Length 400;
 Best Local Similarity 75.0%; Pred. No. 14;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 3 TAAAGCTTAGCATTCTAAAGTCATCATCC 34
 DB 251 TATGCTTAGCAAGCAATATCATCTGTGAGCC 282

RESULT 12

V78616
 ID V78616 standard; DNA; 344 BP.
 AC V78616;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #4305.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 CS Staphylococcus aureus.
 PN EP-786519-N2.
 PD 30-JUL-1997.
 PR 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DT WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1; Page 2911; 3271bp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 344 BP; 92 A; 60 C; 73 G; 116 T;

Query Match 54.9%; Score 19.2; DB 1; Length 344;
 Best Local Similarity 75.0%; Pred. No. 13;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 3 TAAAGCTTAGCATTCTAAAGTCATCATCC 34
 DB 31 TATGCTTAGCAAGCAATATCATCTGTGAGCC 62

RESULT 13

V74342/C
 ID V74342 standard; DNA; 13856 BP.
 AC V74342;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #31.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 CS Staphylococcus aureus.
 PN Key
 FH Location/Qualifiers
 FT misc_feature 1..60
 FT /tag= a
 FT /note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence.
 FT misc-feature
 FT 1801. 1860
 FT /tag- b
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 3601. 3660
 FT /tag- c
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 5401. 5460
 FT /tag- d
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 7201. 7260
 FT /tag- e
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 9001. 9060
 FT /tag- f
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 10801. 10860
 FT /tag- g
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 12601. 12660
 FT /tag- h
 FT "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc-feature
 FT 30-JUL-1997.
 FT 07-JAN-1997: 100117.
 FT 05-JAN-1996: US-009861.
 FT (HUMA-) HUMAN GENOME SCI INC.
 FT Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 FT Rosen CA,
 FT WPI: 97-374922/35.
 FT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 FT stored on computer readable medium and used in the production of
 FT anti-S. aureus vaccines
 FT Claim 1, Page 308-315, 3271pp: English.
 FT This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 FT of the invention. The DNA sequences are recorded on a computer readable
 FT medium, preferably selected from a floppy or hard disk, random access
 FT memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 FT the S. aureus DNA sequences allows putative functions to be assigned so
 FT that protein encoding or regulatory regions of commercial, therapeutic or
 FT industrial importance can be obtained. Specifically, sequences which are
 FT likely to encode antigens have been identified and these polypeptides can
 FT be used in a vaccine composition against S. aureus infection. The
 FT polypeptides can also be used in a kit for the immunodetection of
 FT S. aureus in a sample. S. aureus is implicated in numerous human diseases,
 FT including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 FT skin and surgical wound infections, scalded skin syndrome, toxic shock
 FT syndrome, etc. Organisms transformed with the DNA sequences can be used
 FT for recombinant production of the polypeptides. The new DNA sequences
 FT (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S. aureus DNA sequences contained on the
 CC computer readable medium.
 CC Sequence 13856 BP; 4218 A; 2598 C; 2085 G; 4467 T;
 SQ
 Query Match 54.3%; Score 19; DB 1; Length 13856;
 Best Local Similarity 75.0%; Pred. No. 21;
 Matches 24; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
 QY 3 TAAGCTTAGGCATCTTAAGTCACTATCATCC 34
 DB 8125 TATGCTTAGGCACACTAATCATGTGCACCC 8094
 || ||||| |||| |||| |||| ||
 RESULT 14
 X13998/C
 ID X13998 standard; DNA; 2821 BP.
 AC X13998:
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHPD 984 gene.
 KW GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
 OS peptic ulcer disease; ss.
 OS Helicobacter pylori.
 FH Key location/Qualifiers
 FT CDS 58..2772
 FT /tag- a
 FT W09843478-A1.
 PD 08-OCT-1998.
 PF 01-APR-1998; U06371.
 PR 29-JUL-1997: US-902615.
 PR 01-APR-1997: US-833457.
 PR 24-JUN-1997: US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PA Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR WPI: 98-542293/46.
 DR P-PSDB: W98279.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 1; Page 317-321, 2054pp: English.
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H. pylori GHPD protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these
 CC infections, including acute, chronic, and atrophic gastritis, and peptic
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
 CC for the production of antibodies. The products can also be used for
 CC detection and diagnosis
 CC Sequence 2821 BP; 918 A; 480 C; 663 G; 760 T;
 SQ
 Query Match 54.3%; Score 19; DB 1; Length 2821;
 Best Local Similarity 81.5%; Pred. No. 21;
 Matches 22; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 6 GCTTAGGCATCTTAAGTCACTATCAT 32
 DB 2118 GCTTAGGTCTTAAGTCTCAACAT 2092
 ||||| ||||| ||||| ||||| |||||
 RESULT 15
 T24992
 ID T24992 standard; cDNA to mRNA; 240 BP.
 AC T24992:
 DT 07-NOV-1996 (first entry)
 DE Human gene signature H06S07114.
 KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;
 KW human; cloning; mapping; non-biased library; diagnosis; detection;
 OS cell typing; abnormal cell function; ss.
 OS Homo sapiens.
 PN W09514772-A1.
 PD 01-JUN-1995.

PF 11-NOV-1994: J01916.
 PR 12-NOV-1993: JP-355504.
 PA (MATS/) MATSUBARA K.
 PA (OKUB/) OKUBO K.
 FI Matsubara K, Okubo K;
 DR WPI: 95-206931/27.
 PT Identifying gene signatures in 3'-directed human cDNA library - e.g.
 PT for diagnosis of abnormal cell function, by preparing cDNA that
 PT reflects relative abundance of corresp. mRNA in specific human
 PT tissues
 PS Claim 1: Page 1744; 2245pp; Japanese.
 CC A single-stranded DNA (or its complementary strand or the corresp.
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences
 CC given in T19001-T26837 and which is able to hybridise to part of
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (gene Signature)
 CC sequences were obtained from 3'-directed cDNA libraries prepared
 CC from various human tissues; synthesis of cDNA was initiated from the
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
 CC untranslated sequence is unique to a particular mRNA species, almost
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
 CC is constructed so as to reflect accurately the relative abundance of
 CC different mRNAs in the particular tissue from which it was derived.
 CC The appearance frequency of a given GS in a cDNA library can be
 CC determined (esp. using primers and probes derived from the GS
 CC sequences) as a means of diagnosing abnormal cell function or for
 CC recognising different cell types.
 SQ Sequence 240 BP; 73 A; 41 C; 49 G; 72 T;

Query Match

53.7% Score 18.8; DB 1; Length 240;

Best Local Similarity 76.7%; Pred. No. 19;
 Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TTAAGCTTAGGCATCTTAAGTCACTATCA 31
 ||| ||| ||| ||| ||| ||| ||| |||
 Db 132 TTATGCTTAGACATTTTAAGTAATTATGA 161

Search completed: September 13, 1999, 15:59:31
 Job time: 4798 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:55 ; Search time 1694.61 Seconds

(without alignments)
40.740 Million cell updates/sec

Title: US-09-325-095-17

Perfect score: 35
Sequence: 1 TTTAAGCTTAGGCACTTCTTAAGTCATCATCATCCC 35

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: *
1: em_est11: *
2: em_est12: *
3: em_est13: *
4: em_est14: *
5: em_est15: *
6: em_est16: *
7: em_est17: *
8: em_est18: *
9: em_est19: *
10: em_est20: *
11: em_est21: *
12: em_est22: *
13: em_est23: *
14: em_est24: *
15: em_est25: *
16: em_est26: *
17: em_est27: *
18: em_est28: *
19: em_est29: *
20: em_est30: *
21: em_est31: *
22: em_est32: *
23: em_est33: *
24: em_est34: *
25: em_est35: *
26: em_est36: *
27: em_est37: *
28: em_est38: *
29: em_est39: *
30: em_est40: *
31: em_est41: *
32: em_est42: *
33: em_est43: *
34: em_est44: *
35: em_est45: *
36: em_est46: *
37: em_est47: *
38: em_est48: *
39: em_est49: *
40: em_est50: *
41: em_est51: *
42: em_est52: *
43: em_est53: *
44: em_est54: *
45: em_est55: *
46: em_est56: *
47: em_est57: *
48: em_est58: *
49: em_est59: *
50: em_est60: *
51: em_est61: *
52: em_est62: *
53: em_est63: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	24.4	69.7	417	29	AA190409	AA190409 zp89c10.r
2	20.8	59.4	307	36	AA610665	AA610665 np94a03.s
3	20.4	58.3	593	39	AA887564	AA887564 q141c12.s
4	20.4	58.3	389	44	A1265938	A1265938 q930f06.x
5	20.2	57.7	541	48	A1543543	A1543543 SD10336.5
6	19.8	56.6	314	20	A1161953	A1161953 HSAADJZE H
7	19.8	56.6	595	43	A1161953	A1161953 A010P20U
8	19.8	56.6	557	43	A1162696	A1162696 A022P09U
9	19.8	56.6	491	43	A1165066	A1165066 A074P55u
10	19.8	56.6	842	43	A1188280	A1188280 qd11h07.x
11	19.6	56.0	586	35	Z99424	Z99424 HS299424 DK
12	19.6	56.0	605	41	A1050466	A1050466 uc86e10.y
13	19.6	56.0	749	44	A1314059	A1314059 u125b03.x
14	19.6	56.0	620	53	HSN002526	A1038186 Homo sapi
15	19.6	56.0	762	53	HSN002976	A1038500 Homo sapi
16	19.4	55.4	372	20	T26601	T26601 AB32D7F In
17	19.2	54.9	442	25	N91396	N91396 za16b02.r1
18	19.2	54.9	460	27	AA015969	AA015969 ze30f12.r
19	19.2	54.9	503	30	AA197781	AA197781 mv03f10.r
20	19.2	54.9	390	34	AA528623	AA528623 nfo2e09.s
21	19.2	54.9	133	41	AU006843	AU006843 AU006843
22	19.2	54.9	633	47	A1518990	A1518990 GM44169.3
23	19.2	54.9	422	49	A1628722	A1628722 LY78b01.x
24	19.2	54.9	515	22	R56494	R56494 Y994a06.r1
25	19.2	54.3	427	24	N29688	N29688 YW78a07.s1
26	19.2	54.3	508	24	N29717	N29717 YW78f05.s1
27	19.2	54.3	597	26	W73757	W73757 zd50b08.s1
28	19.2	54.3	493	30	AA242870	AA242870 z164b03.r
29	19.2	54.3	365	33	AA414436	AA414436 v08f10.s
30	19.2	54.3	418	33	AA416936	AA416936 z194b04.s
31	19.2	54.3	464	40	AA905139	AA905139 o113g01.s
32	19.2	54.3	363	40	AA905280	AA905280 o196b01.s
33	19.2	54.3	359	41	A1061048	A1061048 an22e06.x
34	19.2	54.3	349	43	A1220668	A1220668 q901e04.x
35	19.2	54.3	268	47	A1467067	A1467067 v08f10.x
36	19.2	54.3	376	47	A1509866	A1509866 ms30e06.y
37	18.8	53.7	273	21	F03397	F03397 HSC1WA022 n
38	18.8	53.7	461	22	H07908	H07908 Y186h07.s1
39	18.8	53.7	401	22	R44076	R44076 Y926c11.s1
40	18.8	53.7	390	22	R52668	R52668 Y999b02.s1
41	18.8	53.7	483	24	H84045	H84045 YV88f11.s1
42	18.8	53.7	371	24	N33090	N33090 YV33f12.s1
43	18.8	53.7	188	27	AA001067	AA001067 ze48b11.s
44	18.8	53.7	613	27	AA001451	AA001451 ze45g07.s
45	18.8	53.7	471	36	AA620375	AA620375 ae57a05.s

ALIGNMENTS

RESULT 1
LOCUS AA190409/c
DEFINITION zp89c10.r1 Stratiogene Hela cell s3 937216 Homo sapiens cDNA clone
IMAGE:27378 5' similar to SW:P1A_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
RNA sequence.
ACCESSION AA190409


```

/clone_lib="SD Drosophila melanogaster Schneider L2 cell
culture port2"
/lab_host="DHS-alpha"
/Note="Vector: port2; Site:1: EcoRI; Site:2: XhoI; sized
fractionated cDNAs were directly ligated into port2.
Plasmid cDNA library."
BASE COUNT      176 a      105 c      118 g      142 t
ORIGIN

Query Match      57.7%; Score 20.2; DB 48; Length 541;
Best Local Similarity 75.8%; Pred. No. 90;
Matches 25; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      1 TTTAGCTTAGGCACTTCTAAAGTCACTATCATC 33
      1 ||||| ||||| ||||| || |||||
Db      506 TATAGTTTAGGTTTCTTAATTCACCATCAGC 474

RESULT 6
221049/c      314 bp      mRNA      EST      07-FEB-1995
LOCUS      HSAADJZE H, Human adult Brain Cortex tissue Homo sapiens CDNA
DEFINITION similar to (Genbank M95258) human chromosome 4 sequence-tagged site
            sts4-134, mRNA sequence.
ACCESSION 221049
VERSION 927792
KEYWORDS 221049.1 GI:27792
SOURCE EST.
ORGANISM human.
          Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 314)
AUTHORS MRC Human Genome Mapping Project Resource Centre.
TITLE The UK-HGMP CDNA Program
JOURNAL Unpublished (1993)
COMMENT Contact: MRC Human Genome Mapping Project Resource Centre
         Clinical Research Centre
         Watford Road, Harrow, Middlesex HA1 3UD, U.K.
         Email: biocnlp@hgmmp.mrc.ac.uk
         *single read.

FEATURES
  source      Location/Qualifiers
            1..314
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="H, Human adult Brain Cortex tissue"
            /note="Vector: gtl1; clone library-H, Human adult Brain
            Cortex tissue; Cloning vector is gtl1."
BASE COUNT      96 a      65 c      58 g      95 t
ORIGIN

Query Match      56.6%; Score 19.8; DB 20; Length 314;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      2 TTAAGCTTAGGCACTTCTAAAGTCACTATCAT 32
      1 ||||| ||||| ||||| || |||||
Db      138 TAAAGCTTAGGCTTCTCAAAACACATTTTCA 108

RESULT 7
A1161953      595 bp      mRNA      EST      03-DEC-1998
LOCUS      A010P20U Hybrid aspen plasmid library Populus tremula x Populus
DEFINITION tremuloides cDNA 5', mRNA sequence.
ACCESSION A1161953
VERSION 93853238
KEYWORDS A1161953.1 GI:3853238
SOURCE EST.
Populus tremula x Populus tremuloides.

```

```

ORGANISM
Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurossids I; Malpighiales; Sauriceae; Populus.
1 (bases 1 to 595)
REFERENCE
AUTHORS Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A.,
          Holmberg,A., Amthl,B., Bhalero,R., Larsson,M., Villarroel,R., Van
          Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W.,
          Gustafsson,P., Uhlen,M., Sundberg,B. and Lundberg,J.
          Gene discovery in the wood-forming tissues of poplar: Analysis of
          5,692 expressed sequence tags
          Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
          99007314
COMMENT On Jan 9, 1998 this sequence version replaced gi:937424.

Contact: Sterky F
Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se
PCR Primers
FORWARD: AAAGGGGATGCTGCAAGGCG
BACKWARD: GCTCCGGCTGTATCTTGTG
Seq primer: CATTGTAAAGCAGCGCCAG
High quality sequence stop: 595.
Location/Qualifiers
1..595
/organism="Populus tremula x Populus tremuloides"
/db_xref="taxon:47664"
/clone_lib="Hybrid aspen plasmid library"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/note="Vector: pBluescript SK; Site:1: SalI; Site:2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."
BASE COUNT      169 a      120 c      139 g      163 t      4 others
ORIGIN

Query Match      56.6%; Score 19.8; DB 43; Length 595;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      3 TAAAGCTTAGGCACTTCTAAAGTCACTATCATC 33
      1 ||||| ||||| ||||| || |||||
Db      303 TAAAGCTTAGGCTTCTCAAAACACATTTTCA 273

RESULT 8
A1162696      557 bp      mRNA      EST      03-DEC-1998
LOCUS      A022P09U Hybrid aspen plasmid library Populus tremula x Populus
DEFINITION tremuloides cDNA 5', mRNA sequence.
ACCESSION A1162696
VERSION 93853981
KEYWORDS A1162696.1 GI:3853981
SOURCE EST.
Populus tremula x Populus tremuloides.
Populus tremula x Populus tremuloides
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurossids I; Malpighiales; Sauriceae; Populus.
1 (bases 1 to 557)
REFERENCE
AUTHORS Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A.,
          Holmberg,A., Amthl,B., Bhalero,R., Larsson,M., Villarroel,R., Van
          Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W.,

```

TITLE Gustafsson, P., Uhlen, M., Sundberg, B. and Lundberg, J.
Gene discovery in the wood-forming tissues of Poplar: Analysis of
5,692 expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE # 99007314
COMMENT On Sep 12, 1996 this sequence version replaced gl:1398120.

Contact: Sterky, F
Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se

PCR Primers
FORWARD: AAAGGGGATGCTGCGAAGCG
BACKWARD: GCTCCGCTCGTATGTGTG
Seq primer: CGTGTAAACGACGCGCAG
High quality sequence stop: 557.

FEATURES

Location/Qualifiers

1..557
/organism="Populus tremula x Populus tremuloides"
/db_xref="taxon:47664"
/clone_lib="Hybrid aspen plasmid library"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

BASE COUNT 164 a 107 c 125 g 156 t 5 others
ORIGIN

Query Match 56.6%; Score 19.8; DB 43; Length 557;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TAAGCTTAGCATCTTAAGTCACATATC 33
||||| ||||| ||||| ||||| |||||
Db 309 TAAGCAAGCGCTTTTAAAGTCAGTACATC 279

RESULT 9
LOCUS A1165066 491 bp mRNA EST 03-DEC-1998
DEFINITION A074p5u Hybrid aspen plasmid library Populus tremula x Populus
tremuloides cDNA 5', mRNA sequence.
ACCESSION A1165066
NID 93856351
VERSION A1165066.1 GI:3856351
KEYWORDS EST.

REFERENCE
AUTHORS Populus tremula x Populus tremuloides.
Populus tremula x Populus tremuloides
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
euphyllophytes: Spermatophyta: Magnoliophyta: eudicotyledons: core
eudicots: Rosidae: eurosids I: Malpighiales: Salicaceae: Populus.

1 (bases 1 to 491)
Sterky, F., Regan, S., Karlsson, J., Hertzberg, M., Rohde, A.,
Holmberg, A., Amin, B., Bhalerao, R., Larsson, M., Villarroel, R., Van
Montagu, M., Sandberg, G., Olsson, O., Teerl, T.T., Boerjan, W.,
Gustafsson, P., Uhlen, M., Sundberg, B. and Lundberg, J.

TITLE Gene discovery in the wood-forming tissues of Poplar: Analysis of
5,692 expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
MEDLINE 99007314
COMMENT On Apr 18, 1995 this sequence version replaced gl:775315.
Contact: Sterky, F

Department of Biotechnology
Royal Institute of Technology (KTH)
Teknikringen 34, S-100 44 STOCKHOLM, Sweden
Tel: +46 8 790 8287
Fax: +46 8 24 54 52
Email: fredrik@biochem.kth.se

PCR Primers
FORWARD: AAAGGGGATGCTGCGAAGCG
BACKWARD: GCTCCGCTCGTATGTGTG
Seq primer: CGTGTAAACGACGCGCAG
High quality sequence stop: 491.

FEATURES

Location/Qualifiers

1..491
/organism="Populus tremula x Populus tremuloides"
/db_xref="taxon:47664"
/map="7"
/clone_lib="Hybrid aspen plasmid library"
/tissue_type="Cambial region"
/dev_stage="1.5 m actively growing tree"
/lab_host="E.coli"
/note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
Cambial region tissues, including developing xylem, the
meristematic cambial zone and the developing and mature
phloem, was harvested from 1.5 m actively growing trees.
cDNA was prepared and cloned into lambda gt22a. DNA was
isolated and subcloned into pBluescript SK using SalI and
NotI restriction enzymes."

BASE COUNT 146 a 98 c 109 g 137 t 1 others
ORIGIN

Query Match 56.6%; Score 19.8; DB 43; Length 491;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 3 TAAGCTTAGCATCTTAAGTCACATATC 33
||||| ||||| ||||| ||||| |||||
Db 274 TAAGCAAGCGCTTTTAAAGTCAGTACATC 244

RESULT 10
LOCUS A1188280 842 bp mRNA EST 13-OCT-1998
DEFINITION qd11h07.x1 Soares-Placenta_8c9weeks_2MDHPeto9M Homo sapiens CDNA
clone IMAGE:1723453 3' similar to TR:Q13539 Q13539 MARINER
TRANSPOSASE. ;, mRNA sequence.
ACCESSION A1188280
NID 93739489
VERSION A1188280.1 GI:3739489
KEYWORDS EST.

REFERENCE
AUTHORS Homo sapiens
human.
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;
Eutheria: Primates: Catarrhini: Hominiidae: Homo.

1 (bases 1 to 842)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Aug 21, 1998 this sequence version replaced.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -400P from GIDCO
High quality sequence stop: 464.

FEATURES

Location/Qualifiers

1..842
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1723453"

/clone.lib="Soares.placenta.8to9weeks.2NHP8to9w"
 /dev_stage="two placenta: one from 8 weeks and another
 from 9 weeks post conception"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: placenta; Vector: pT73D (Pharmacia) with a
 modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCACTGAAAGTGGAGCGCGGAGATTTTATTTT 3'],
 T7ACCACTGAAAGTGGAGCGCGGAGATTTTATTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bonaldo."

BASE COUNT 228 a 180 c 152 g 279 t 3 others
 ORIGIN

Query Match 56.6%; Score 19.6; DB 43; Length 842;
 Best Local Similarity 77.4%; Pred. No. 1.3e+02;
 Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TTAAGCTTAGCATCTTAAGTCATCATCATC 32
 1 ||||| 111 ||||| 111 |||||
 DB 349 TAAAGCTTGGCTTTCGCAAAACACTCATC 379

RESULT 11
 299424 586 bp mRNA EST 22-SEP-1997
 LOCUS HS299424 DKFZphamyl Homo sapiens cDNA clone DKFZphamyl_1d3, mRNA
 DEFINITION sequence.
 ACCESSION 299424
 MID 92415664
 VERSION 299424.1 GI:2415664
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 586)
 Kohn, B., Wiemann, S., Ebert, L. and Poustka, A.
 TITLE Human ESTs (Kohn, B. et al.)
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393514.
 Contact: B. Kohn
 Institution Molekulare Genomanalyse
 Deutsches Krebsforschungszentrum Heidelberg
 Im Neuenheimer Feld 506, D-69120 Heidelberg, FRG.

FEATURES
 source
 1..586
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone.lib="DKFZphamyl_1d3"
 /clone.lib="DKFZphamyl"
 /issue_type="amygdala"
 /dev_stage="adult"

BASE COUNT 184 a 110 c 136 g 156 t
 ORIGIN

Query Match 56.0%; Score 19.6; DB 35; Length 586;
 Best Local Similarity 73.5%; Pred. No. 1.6e+02;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TTAAGCTTAGCATCTTAAGTCATCATCATC 35
 1 ||||| 111 ||||| 111 |||||
 DB 512 TGAAGCTTGGCTTGAAGTGAATATCATCC 545

RESULT 12
 AI050466 605 bp mRNA EST 09-JUL-1998

DEFINITION uc06e10.y1 Sugano mouse kidney mKia Mus musculus cDNA clone
 IMAGE:1432554 5' similar to SW:CI0H_HUMAN Q16181 CDC10 PROTEIN
 HOMOLOG.; mRNA sequence.

ACCESSION AI050466
 MID 93299583
 VERSION AI050466.1 GI:3299583
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 605)
 Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Scheinberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Jan 17, 1998 this sequence version replaced gi:2044433.

Contact: Maria M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LBL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:916622
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 349.

FEATURES
 source
 1..605
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /map="14"
 /clone.lib="IMAGE:1432554"
 /clone.lib="Sugano mouse kidney mKia"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
 (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGTGGCTTTTATTTTATTTTATTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTGGCTTACTGG], digested
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTTCCTCTTAAGCTGTGG and 3' end
 primer CGACCTGCAGCTCGAGACA."

BASE COUNT 210 a 109 c 124 g 162 t
 ORIGIN

Query Match 56.0%; Score 19.6; DB 41; Length 605;
 Best Local Similarity 73.5%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 TTAAGCTTAGCATCTTAAGTCATCATCATC 35
 1 ||||| 111 ||||| 111 |||||
 DB 421 TGAAGCTTGGCTTGAAGTGAATATCATCC 454

RESULT 13
 ATJ14059/C

LOCUS A1314059 749 bp mRNA EST 17-DEC-1998
 DEFINITION u75b03.x1 Sugano mouse kidney mK1a Mus musculus cDNA clone
 IMAGE:1920941 3' similar to SM:C10H_HUMAN Q16181 CDC10 PROTEIN
 HOMOLOG. ;, mRNA sequence.
 ACCESSION A1314059
 MID g4029269
 VERSION A1314059.1 GI:4029269
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 749)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Jan 14, 1998 this sequence version replaced gi:1798001.
 TITLE
 JOURNAL
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (Info@image.lnl.gov) for further information.
 MGI:977233
 Possible reversed clone: similarity on wrong strand
 Seq primer: custom primer used
 High quality sequence stop: 431.
 Location/Qualifiers
 1..749
 /organism="Mus musculus"
 /strain="C57BL"
 /db_xref="taxon:10090"
 /map="21922"
 /clone="IMAGE:1920941"
 /clone_lib="Sugano mouse kidney mK1a"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: pME18S-FL3; Site1: DraIII
 (CACGCTGTG); Site2: DraIII (CACCATGTG); 1st strand cDNA
 was primed with an oligo(dT) primer
 [ATGTGGCCCTTTTCTTTTCTTTT]; double-stranded cDNA was
 ligated to a DraIII adaptor [TGTGCGCTACTG], digested
 and cloned into distinct DraIII sites of the pME18S-FL3
 vector (5' site CACGCTGTG, 3' site CACCATGTG). XhoI should
 be used to isolate the cDNA insert. Size selection was
 performed to exclude fragments <1.5kb. Library
 constructed by Dr. Sumio Sugano (University of Tokyo
 Institute of Medical Science). Custom primers for
 sequencing: 5' end primer CTCCTGCTCTTAAAGCTGGC and 3' end
 primer CGACCTGACGCTGAGACA."
 BASE COUNT 160 a 179 c 112 g 294 t 4 others
 ORIGIN
 Query Match 56.0%; Score 19.6; DB 44; Length 749;
 Best Local Similarity 73.5%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Oy 2 TTAAGCTTAGGCATTTCTAAAGTCATCATCATCC 35
 Db 680 TGAACGTTGGCATGAAAAAGTGAATATCATCC 647
 RESULT :14

HSN002526
 ID HSN002526 standard; RNA; EST; 620 BP.
 AC AL038186;
 XX AL038186.1
 SV e1396521
 NI e1396521
 DT 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 DE Homo sapiens mRNA: EST DKFZ566N1624_r1 (from clone
 DKFZ566N1624)
 EST; expressed sequence tag.
 OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 OC Primates; Catarrhini; Homiidae; Homo.
 XX [1]
 RA Bloeker H., Boecker M., Brandt P., Mewes W., Gassenhuber J.,
 Wiemann S.;
 RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
 RL MIPS, Am Kiofersplitz 18a D-82152 Martinsried, GERMANY
 CC Clone from S. Wiemann, sequenced by GBF within the cDNA
 CC sequencing consortium of the German Genome Project
 CC s1 sequence is also available
 CC This clone is available at the RZPD in Berlin
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
 CC
 XX
 FH Key Location/Qualifiers
 FT source 1..620
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="DKFZ566N1624"
 FT /clone_lib="566 (synonym: hKd2). Vector pAMP1; host
 FT X1-ZBlue; sites NotI + SalI"
 FT /dev_stage="fetal"
 FT /tissue_type="kidney"
 FT
 XX
 SO Sequence 620 BP; 196 A; 116 C; 145 G; 163 T; 0 other;
 Query Match 56.0%; Score 19.6; DB 53; Length 620;
 Best Local Similarity 73.5%; Pred. No. 1.5e+02;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
 Oy 2 TTAAGCTTAGGCATTTCTAAAGTCATCATCATCC 35
 Db 533 TGAACGTTGGCATGAAAAAGTGAATATCATCC 566
 RESULT :15
 HSN002976
 ID HSN002976 standard; RNA; EST; 762 BP.
 AC AL038500;
 XX AL038500.1
 SV e1396835
 NI e1396835
 DT 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 DE Homo sapiens mRNA: EST DKFZ566D1246_r1 (from clone

```

DE DKEZP566D1246)
XX
XX EST: expressed sequence tag.
XX
XX
XX Homo sapiens (human)
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC Primates; Catarrhini; Homnidae; Homo.
XX
XX [1]
RN 1-762
RP Ottenwaelder B., Obermaier B., Mewes W., Gassenhuber J., Wiemann S.;
RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX
XX Clone from S. Wiemann, sequenced by Medigenomix within the CDNA
CC sequencing consortium of the German Genome Project
CC s1 sequence also available
CC This clone is available at the RZPD in Berlin
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX
XX
XX Key Location/Qualifiers
FH 1..762
FT source
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKFZP566D1246"
FT /clone_lib="566 (synonym: hfk42). Vector pAMP1; host
FT X1-2blue; sites NotI + SalI"
FT /dev_stage="fetal"
FT /tissue_type="kidney"
XX
SQ Sequence 762 bp; 262 A; 125 C; 164 G; 211 T; 0 other;

```

```

Query Match 56.0%; Score 19.6; DB 53; Length 762;
Best Local Similarity 73.5%; Pred. No. 1.5e+02;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 2 TTAAGCTTAGCATCTTAAGTCAATCATCC 35
| |||| | |||| | |||| | |||| |
DB 396 TGAAGCCTTGCAATGAAAGTGAATATCATCC 429

```

Search completed: September 13, 1999, 15:45:56
 Job time: 4044 sec

FEATURES
source
Location/Qualifiers
1..638
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/cell_type="oocyte"
107..583
/gene="ubc9"
107..583
/gene="ubc9"
/note="ubc9"
/codon_start=1
/product="E2 ubiquitin conjugating enzyme"
/protein_id="AA57336.1"
/db_xref="PID:2078331"
/db_xref="GI:2078331"
/translation="MSGIALSLAQRKAKRDPGFVAVPTKNDGTMNLMNMECA
IPGKGTPEGGGLFRLMLFKDYPSPSPKPEPLFPHNVPSTGVCSTIEEDKD
WRPATIKILIGIOELNPNPDPAQAEAVTIYQGNVEYERKRAQAKKFAFS"
BASE COUNT 214 a 131 c 147 g 146 t
ORIGIN

Query Match 100.0%; Score 18; DB 4; Length 638;
Best Local Similarity 100.0%; Pred. No. 6.1; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
DB 638 GACTCGAGTCGACATCGA 621

RESULT 2
A24382 35 bp DNA PAT 21-MAR-1995
LOCUS A24382 35 bp DNA
DEFINITION d117 adaptor primer.
ACCESSION A24382
NID 9904477
VERSION A24382.1 GI:904477
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 35)
AUTHORS Samlencos, P., De Taxis du Poet, P., Mittl, G. and Seacherl, E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: EP 0501821-A 22 02-SEP-1992;
FARMITALIA CARLO ERBA S.r.L.
FEATURES Location/Qualifiers
source 1..35
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
DB 1 GACTCGAGTCGACATCGA 18

RESULT 3
A27645 35 bp DNA PAT 04-JUN-1995
LOCUS A27645 35 bp DNA
DEFINITION Hybrid d117-adaptor primer.
ACCESSION A27645
NID 91248481
VERSION A27645.1 GI:1248481
KEYWORDS
SOURCE synthetic construct.

ORGANISM synthetic construct
artificial sequence.
REFERENCE 1 (bases 1 to 35)
AUTHORS POLYPHENOL OXIDASE GENES
JOURNAL Patent: WO 9302195-A 1 04-FEB-1993;
FEATURES Location/Qualifiers
source 1..35
/organism="synthetic construct"
/db_xref="taxon:32630"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
DB 1 GACTCGAGTCGACATCGA 18

RESULT 4
A37244 35 bp DNA PAT 05-MAR-1997
LOCUS A37244 35 bp DNA
DEFINITION Sequence 13 from Patent WO9403609.
ACCESSION A37244
NID 92294355
VERSION A37244.1 GI:2294355
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 35)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 13 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES Location/Qualifiers
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7; Mismatches 0; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
DB 1 GACTCGAGTCGACATCGA 18

RESULT 5
A37246 18 bp DNA PAT 05-MAR-1997
LOCUS A37246 18 bp DNA
DEFINITION Sequence 15 from Patent WO9403609.
ACCESSION A37246
NID 92294357
VERSION A37246.1 GI:2294357
KEYWORDS
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 18)
AUTHORS Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
TITLE EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM

JOURNAL Patent: WO 9403609-A 15 17-FEB-1994;
IMP CANCER RES TECH (GB)
COMMENT Other publication JP 8503124T 960409.
FEATURES Location/Qualifiers
Source 1. .18
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGACTCGACATCGA 18
|||||

Db 1 GACTCGACTCGACATCGA 18

RESULT 6
LOCUS A40372 18 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 28 from Patent WO9425489.
ACCESSION A40372
NID 92296421
VERSION A40372.1 GI:2296421
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Mohapatra,S.S. and Sehon,A.H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL Patent: WO 9425489-A 28 10-NOV-1994;
UNIV MANITOBA (CA)
COMMENT Other publication AU 6674094 941121.
FEATURES Location/Qualifiers
Source 1. .18
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 3 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 18;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGACTCGACATCGA 18
|||||

Db 1 GACTCGACTCGACATCGA 18

RESULT 7
LOCUS A40373 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 29 from Patent WO9425489.
ACCESSION A40373
NID 92296422
VERSION A40373.1 GI:2296422
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Mohapatra,S.S. and Sehon,A.H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL Patent: WO 9425489-A 29 10-NOV-1994;
UNIV MANITOBA (CA)
COMMENT Other publication AU 6674094 941121.
FEATURES Location/Qualifiers
Source 1. .35

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGACTCGACATCGA 18
|||||

Db 1 GACTCGACTCGACATCGA 18

RESULT 8
LOCUS A42335 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 7 from Patent WO9502057.
ACCESSION A42335
NID 92297812
VERSION A42335.1 GI:2297812
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
Kamali,T., Page,M.J. and Spence,P.
TITLE PROTEIN TYROSINE KINASE AND LIGANDS THEREOF
JOURNAL Patent: WO 9502057-A 7 19-JAN-1995;
CANCER RES INST (GB)
COMMENT Other publication AU 7080994 950206.
FEATURES Location/Qualifiers
Source 1. .35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGACTCGACATCGA 18
|||||

Db 1 GACTCGACTCGACATCGA 18

RESULT 9
LOCUS A42384 35 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 7 from Patent WO9502187.
ACCESSION A42384
NID 92297858
VERSION A42384.1 GI:2297858
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Gusterson,B.A., Crompton,M.R., Mitchell,P.J., Barker,K.T.,
Martindale,J.E., Page,M.J. and Spence,P.
TITLE CELL GROWTH FACTOR RECEPTORS
JOURNAL Patent: WO 9502187-A 7 19-JAN-1995;
CANCER RES INST (GB)
COMMENT Other publication AU 7081094 950206.
FEATURES Location/Qualifiers
Source 1. .35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 10

LOCUS A46467 35 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 5 from Patent WO9526402.
ACCESSION A46467
NID 92300644
VERSION A46467.1 GI:2300644
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 35)
AUTHORS Knox, D. P., Smith, S. K., Smith, W. D., Redmond, D. and Murray, J.
TITLE VACCINES AGAINST HELMINTHIC PARASITES
JOURNAL Patent: WO 9526402-A 5 05-OCT-1995;
MALDINCKRODT VETERINARY INC (US)
Other Publication ZA 2182178 951005
Other Publication AU 1956495 951017.
FEATURES
source 1. .35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 11

LOCUS A59198 40 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 7 from Patent WO9704108.
ACCESSION A59198
NID 93714579
VERSION A59198.1 GI:3714579
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 40)
AUTHORS Schuster, E., Spiessler, B., Tltze, K., Gottschalk, M., Khan, N. Q.,
Wolf, S. and Plainer, H.
TITLE LEUCINE AMINOPEPTIDASES PRODUCED RECOMBINANTLY FROM ASPERGILLUS
JOURNAL Patent: WO 9704108-A 7 06-FEB-1997;
ROEHM GMBH (DE)
Other Publication DE 19526485 970123.
FEATURES
source 1. .40
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 24 t 1 others
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 40;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 12

LOCUS A59199 20 bp DNA PAT 06-MAR-1998
DEFINITION Sequence 8 from Patent WO9704108.
ACCESSION A59199
NID 93714580
VERSION A59199.1 GI:3714580
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 20)
AUTHORS Schuster, E., Spiessler, B., Tltze, K., Gottschalk, M., Khan, N. Q.,
Wolf, S. and Plainer, H.
TITLE LEUCINE AMINOPEPTIDASES PRODUCED RECOMBINANTLY FROM ASPERGILLUS
JOURNAL Patent: WO 9704108-A 8 06-FEB-1997;
ROEHM GMBH (DE)
Other Publication DE 19526485 970123.
FEATURES
source 1. .20
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 5 t
ORIGIN

Query Match 100.0%; Score 18; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 6.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 13

LOCUS A68608 35 bp DNA PAT 06-MAY-1999
DEFINITION Sequence 8 from Patent WO9749726.
ACCESSION A68608
NID 94759636
VERSION A68608.1 GI:4759636
KEYWORDS
SOURCE
ORGANISM
REFERENCE 1 (bases 1 to 35)
AUTHORS Mele, A., De, S. R., Parente, D. and Colnaghi, M. I.
TITLE RECOMBINANT RIBOSOMAL INHIBITOR PROTEIN (RIP) AND USE AS
JOURNAL IMMUNOCONJUGATE
Patent: WO 9749726-A 31-DEC-1997;
MINISTERO UNI RICERCA SCIENT E (IT)
Other Publication IT FI960155 19971229.
FEATURES
source 1. .35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Job time: 4551 sec

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 14

AR012367 AR012367 35 bp DNA PAT 04-DEC-1998
LOCUS DEFINITION Sequence 5 from patent US 5763400.
ACCESSION AR012367
MID 93970357
VERSION AR012367.1 GI:3970357
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 35)
AUTHORS Adams,M.E. and Zltman,D.
TITLE Ecdysis-triggering hormone compositions
JOURNAL Patent: US 5763400-A 5 09-JUN-1998;
FEATURES location/Qualifiers
source 1..35

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN /organism="unknown"

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

RESULT 15

113679 113679 35 bp DNA PAT 08-AUG-1995
LOCUS DEFINITION Sequence 35 from patent US 5439820.
ACCESSION 113679
MID 9996745
VERSION 113679.1 GI:996745
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

REFERENCE 1 (bases 1 to 35)
AUTHORS Sarmientos,P., De Taxis du Poet,P., Nitli,G. and Scacheri,E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: US 5439820-A 35 08-AUG-1995;
FEATURES location/Qualifiers
source 1..35

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN /organism="unknown"

Query Match 100.0%; Score 18; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|||||
Db 1 GACTCGAGTCGACATCGA 18

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

UM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:31 ; Search time 255.05 Seconds

(without alignments)
17.657 Million cell updates/sec

Title: US-09-325-095-18

Perfect score: 18

Sequence: 1 GACTCGAGTCGACATCGA 18

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N.Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	100.0	39	1	Q23200
2	18	100.0	32	1	Q23201
3	18	100.0	22	1	Q23202
4	18	100.0	38	1	Q23203
5	18	100.0	39	1	Q23205
6	18	100.0	39	1	Q28423
7	18	100.0	32	1	Q28424
8	18	100.0	22	1	Q28425
9	18	100.0	38	1	Q31777
10	18	100.0	32	1	Q31778
11	18	100.0	35	1	Q36668
12	18	100.0	18	1	Q48416
13	18	100.0	39	1	Q52547
14	18	100.0	35	1	Q52503
15	18	100.0	35	1	Q59015
16	18	100.0	18	1	Q59017
17	18	100.0	18	1	Q57021
18	18	100.0	18	1	Q57023
19	18	100.0	19	1	Q44401
20	18	100.0	35	1	Q71021
21	18	100.0	35	1	Q63873
22	18	100.0	35	1	Q65594
23	18	100.0	18	1	Q78806
24	18	100.0	35	1	Q78807
25	18	100.0	35	1	Q84786
26	18	100.0	32	1	Q86055
27	18	100.0	38	1	Q86054
28	18	100.0	35	1	Q90167
29	18	100.0	35	1	Q89878
30	18	100.0	35	1	T05321
31	18	100.0	1296	1	T05325
32	18	100.0	35	1	Q94244
33	18	100.0	35	1	T10696
34	18	100.0	19	1	T10697
35	18	100.0	901	1	T10698
36	18	100.0	35	1	T16211
37	18	100.0	35	1	T10276
38	18	100.0	35	1	T03701
39	18	100.0	35	1	T11753
40	18	100.0	18	1	T11754
41	18	100.0	37	1	Q74002
42	18	100.0	27	1	Q74003
43	18	100.0	35	1	T35802

44 18 100.0 35 1 T28797 MMIV reverse trans
45 18 100.0 18 1 V56009 D. discoideum Dpi

ALIGNMENTS

RESULT 1
Q23200 standard; DNA; 39 BP.
ID Q23200:
AC 29-JUL-1992 (first entry)
DE ZC2487 oligo-d(T), XhoI, Sall, ClaI primer for AOH DNA.
KW Acyloxylase, septicemia; LPS; vaccine; human; ss.
OS Synthetic.
PN WO920444-A.
PD 19-MAR-1992.
PF 11-SEP-1991; U06569.
PR 12-SEP-1990; US-581342.
PA (ZYMO-) ZYMOGENETICS INC.
PI (TEXA) UNIV OF TEXAS SYSTEM.
DR Ohara PJ, Hagen FS, Grant FJ, Munford RS;
PT WPI; 92-114352/14.
PT Prodn. of acyl:oxy:acyl-hydrolase in eukaryotes - using DNA
constructed encoding whole protein or its sub-units, and used to
PT treat gram negative sepsis.
PS Table 1; Page 29; 97pp; English.
CC The primer was used to synthesize cDNA from U-937 polyA+ RNA.
CC Both 5' and 3' templates were prepd. The resulting cDNAs were
CC enriched and amplified, and were used for the isolation of the
CC complete gene sequence for AOH. The gene was used to express
CC the recombinant enzyme, which can be used prophylactically and
CC therapeutically against gram-negative septicemia in mammals.
CC For treatment, 1 ug-10 mg AOH is given per 70 kg body wt. The
CC recombinant enzyme is produced in higher levels than found
CC naturally in neutrophils and is more easily purified. See also
CC Q23190-215.
SQ Sequence 39 BP: 6 A; 6 C; 6 G; 21 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
Best local similarity 100.0%; Pred. No. 0.33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
DB 1 GACTCGAGTCGACATCGA 18

RESULT 2
Q23201 standard; DNA; 32 BP.
ID Q23201:
AC 29-JUL-1992 (first entry)
DE ZC2488 oligo-d(C), XhoI, Sall, ClaI primer for AOH DNA.
KW Acyloxylase, septicemia; LPS; vaccine; human; ss.
OS Synthetic.
PN WO920444-A.
PD 19-MAR-1992.
PF 11-SEP-1991; U06569.
PR 12-SEP-1990; US-581342.
PA (ZYMO-) ZYMOGENETICS INC.
PI (TEXA) UNIV OF TEXAS SYSTEM.
DR Ohara PJ, Hagen FS, Grant FJ, Munford RS;
PT WPI; 92-114352/14.
PT Prodn. of acyl:oxy:acyl-hydrolase in eukaryotes - using DNA
constructed encoding whole protein or its sub-units, and used to
PT treat gram negative sepsis.
PS Table 1; Page 29; 97pp; English.
CC The primer was used for second strand synthesis of cDNA from U-937
CC polyA+ RNA. Both 5' and 3' templates were prepd. The resulting
CC cDNAs were enriched and amplified, and were used for the isolation
CC of the complete gene sequence for AOH. The gene was used to

CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOA^H is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 32 BP; 6 A; 16 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

RESULT 3
 ID Q23202 standard; DNA: 22 BP.
 AC Q23202;
 DT 29-JUL-1992 (first entry)
 DE ZC2489 primer for AOA^H DNA.
 KM Acyloxycyl hydrolase; septicemia; LPS; vaccine; human; ss.
 OS Synthetic.
 PN WO9204444-A.
 PD 19-MAR-1992.
 PF 11-SEP-1991; U06569.
 PR 12-SEP-1990; US-581342.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI (TEXA-) UNIV OF TEXAS SYSTEM.
 PI Ohara PJ, Hagen FS, Grant FU, Munford RS;
 DR WPI: 92-114352/14.
 PT Prodn. of acyl:oxyl-hydrolase in eukaryotes - using DNA
 PT constructed encoding whole protein or its sub-units, and used to
 PT treat gram negative sepsis.
 PS Table 1; Page 29; 97pp; English.
 CC The primer was used for PCR amplification of cDNA from U-937
 CC polyA+ RNA. The resulting DNA was used for the isolation
 CC of the complete gene sequence for AOA^H. The gene was used to
 CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOA^H is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 22 BP; 6 A; 6 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 22;
 Best Local Similarity 100.0%; Pred. No. 0.32;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

RESULT 4
 ID Q23203 standard; DNA: 38 BP.
 AC Q23203;
 DT 29-JUL-1992 (first entry)
 DE ZC2633 primer for AOA^H DNA.
 KM Acyloxycyl hydrolase; septicemia; LPS; vaccine; human; ss.
 OS Synthetic.
 PN WO9204444-A.
 PD 19-MAR-1992.
 PF 11-SEP-1991; U06569.
 PR 12-SEP-1990; US-581342.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI (TEXA-) UNIV OF TEXAS SYSTEM.

PI Ohara PJ, Hagen FS, Grant FU, Munford RS;
 DR WPI: 92-114352/14.
 PT Prodn. of acyl:oxyl-hydrolase in eukaryotes - using DNA
 PT constructed encoding whole protein or its sub-units, and used to
 PT treat gram negative sepsis.
 PS Table 1; Page 29; 97pp; English.

CC The primer was used for PCR amplification of cDNA from U-937
 CC polyA+ RNA. The resulting DNA was used for the isolation
 CC of the complete gene sequence for AOA^H. The gene was used to
 CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOA^H is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 38 BP; 11 A; 9 C; 12 G; 6 T;

Query Match 100.0%; Score 18; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 17 GACTCGAGTCGACATCGA 34

RESULT 5
 ID Q23205 standard; DNA: 39 BP.
 AC Q23205;
 DT 29-JUL-1992 (first entry)
 DE ZC2632 primer for AOA^H DNA.
 KM Acyloxycyl hydrolase; septicemia; LPS; vaccine; human; ss.
 OS Synthetic.
 PN WO9204444-A.
 PD 19-MAR-1992.
 PF 11-SEP-1991; U06569.
 PR 12-SEP-1990; US-581342.
 PA (ZYMO-) ZYMOGENETICS INC.
 PI (TEXA-) UNIV OF TEXAS SYSTEM.
 PI Ohara PJ, Hagen FS, Grant FU, Munford RS;
 DR WPI: 92-114352/14.
 PT Prodn. of acyl:oxyl-hydrolase in eukaryotes - using DNA
 PT constructed encoding whole protein or its sub-units, and used to
 PT treat gram negative sepsis.
 PS Table 1; Page 29; 97pp; English.
 CC The primer was used for PCR amplification of cDNA from U-937
 CC polyA+ RNA. The resulting DNA was used for the isolation
 CC of the complete gene sequence for AOA^H. The gene was used to
 CC express the recombinant enzyme, which can be used prophylactically
 CC and therapeutically against Gram-negative septicemia in mammals.
 CC For treatment, 1 ug-10 mg AOA^H is given per 70 kg body wt. The
 CC recombinant enzyme is produced in higher levels than found
 CC naturally in neutrophils and is more easily purified. See also
 CC Q23190-215.
 SQ Sequence 39 BP; 13 A; 10 C; 10 G; 6 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 18 GACTCGAGTCGACATCGA 35

RESULT 6
 ID Q28423 standard; cDNA: 39 BP.
 AC Q28423;
 DT 25-FEB-1993 (first entry)

DE Human galanin cDNA PCR primer ZC2487.
KW Polymerase chain reaction; ss.
OS Synthetic.
PN WO9215015-A.
PD 03-SEP-1992.
PF 25-FEB-1992; U01469.
PR 25-FEB-1991; US-662221.
PS 31-DEC-1991; US-816285.
PA (UYBO-) UNIV BOARDS WASHINGTON.
PI (ZYMO-) ZYMOGENETICS INC.
PI Kowalyk S, McKnight GL, Smith RA;
DR WPI: 92-316309/38.
PT Detecting presence of galanin antagonist - by exposing isolated
PT galanin analogue to galanin receptor in presence of natural
PT galanin, used for treating e.g. type II diabetes
PS Example: Page 25; 73pp; English.
CC The sequence is that of a degenerate primer encoding a galanin DNA
CC sequence which was used to amplify galanin cDNA sequences by the
CC polymerase chain reaction (PCR) method. It also contains a 5' tail
CC of 10 nucleotides for subcloning. See also Q28421-Q28432.
SQ Sequence 39 BP; 6 A; 6 C; 6 G; 21 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTCGAGTCGACATCGA 18
DB 1 GACTCGAGTCGACATCGA 18

RESULT 7
Q28424
ID Q28424 standard; cDNA; 32 BP.
AC Q28424:
DT 25-FEB-1993 (first entry)
DE Human galanin cDNA PCR primer ZC2488.
KW Polymerase chain reaction; ss.
OS Synthetic.
PN WO9215015-A.
PD 03-SEP-1992.
PF 25-FEB-1992; U01469.
PR 25-FEB-1991; US-662221.
PS 31-DEC-1991; US-816285.
PA (UYBO-) UNIV BOARDS WASHINGTON.
PI (ZYMO-) ZYMOGENETICS INC.
PI Kowalyk S, McKnight GL, Smith RA;
DR WPI: 92-316309/38.
PT Detecting presence of galanin antagonist - by exposing isolated
PT galanin analogue to galanin receptor in presence of natural
PT galanin, used for treating e.g. type II diabetes
PS Example: Page 25; 73pp; English.
CC The sequence is that of a degenerate primer encoding a galanin DNA
CC sequence which was used to amplify galanin cDNA sequences by the
CC polymerase chain reaction (PCR) method. It also contains a 5' tail
CC of 10 nucleotides for subcloning. See also Q28421-Q28432.
SQ Sequence 32 BP; 6 A; 16 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 32;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
DB 1 GACTCGAGTCGACATCGA 18

RESULT 8
Q28425
ID Q28425 standard; cDNA; 22 BP.
AC Q28425;

DT 25-FEB-1993 (first entry)
DE Human galanin cDNA PCR primer ZC2489.
KW Polymerase chain reaction; ss.
OS Synthetic.
PN WO9215015-A.
PD 03-SEP-1992.
PF 25-FEB-1992; U01469.
PR 25-FEB-1991; US-662221.
PS 31-DEC-1991; US-816285.
PA (UYBO-) UNIV BOARDS WASHINGTON.
PI (ZYMO-) ZYMOGENETICS INC.
PI Kowalyk S, McKnight GL, Smith RA;
DR WPI: 92-316309/38.
PT Detecting presence of galanin antagonist - by exposing isolated
PT galanin analogue to galanin receptor in presence of natural
PT galanin, used for treating e.g. type II diabetes
PS Example: Page 25; 73pp; English.
CC The sequence is that of a degenerate primer encoding a galanin DNA
CC sequence which was used to amplify galanin cDNA sequences by the
CC polymerase chain reaction (PCR) method. It also contains a 5' tail
CC of 10 nucleotides for subcloning. See also Q28421-Q28432.
SQ Sequence 22 BP; 6 A; 6 C; 6 G; 4 T;

Query Match 100.0%; Score 18; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACTCGAGTCGACATCGA 18
DB 1 GACTCGAGTCGACATCGA 18

RESULT 9
Q31777
ID Q31777 standard; DNA; 38 BP.
AC Q31777:
DT 20-APR-1993 (first entry)
DE Human pancreatic islet cell GAD PCR primer ZC2633.
KW Human pancreatic islet cell GAD PCR primer ZC2633.
OS Synthetic.
PN WO9220811-A.
PD 26-NOV-1992.
PF 14-MAY-1992; U04079.
PR 15-MAY-1991; US-702162.
PS (UNIW-) UNIV WASHINGTON.
PA (ZYMO-) ZYMOGENETICS INC.
PI Foster DC, Grublin CE, Hagopian W, Karlsten AE, Lermmark A, Ohara PJ;
DR WPI: 92-415789/50.
PT Polynucleotide encoding human islet glutamic acid decarboxylase -
PT used to test for auto-antibodies against itself to diagnose
PT insulin dependent diabetes mellitus
PS Example: Page 25; 45pp; English.
CC The sequence is that of an oligonucleotide primer ZC2633 which
CC is complementary to the EcoRI adapter and was used in the second
CC PCR step to prime the plus strand. It was used to obtain full
CC length human pancreatic islet GAD cDNA clones.
SQ Sequence 38 BP; 11 A; 9 C; 12 G; 6 T;

Query Match 100.0%; Score 18; DB 1; Length 38;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
DB 17 GACTCGAGTCGACATCGA 34

RESULT 10
Q31778
ID Q31778 standard; DNA; 32 BP.
AC Q31778;

DT 20-APR-1993 (first entry)
 DE Human pancreatic islet cell GAD PCR primer ZC2488.
 KW Polymerase chain reaction; glutamic acid decarboxylase; ss.
 OS Synthetic.
 PN WO9220811-A.
 PD 26-NOV-1992.
 PF 14-MAY-1992; U04079.
 PR 15-MAY-1991; US-702162.
 PA (UNIV) UNIV WASHINGTON.
 PI (ZYMO) ZYMOGENETICS INC.
 DR Foster DC, Grubin CE, Hagopian W, Karlson AE, Lernmark A, Ohara PJ;
 WPI: 93-415789/50.
 PT Polynucleotide encoding human islet glutamic acid decarboxylase -
 used to test for auto-antibodies against itself to diagnose
 PT Insulin dependent diabetes mellitus
 PS Example: Page 25; 45pp; English.
 CC The sequence is that of an oligonucleotide primer ZC2488 with an
 CC EcoRI adapter which was used on second strand synthesis to
 CC generate a cDNA population enriched for glutamic acid decarboxylase
 CC (GAD) but still heterogeneous due to non-specific pairing of the
 CC internal primer. It was used to obtain full-length human pancreatic
 CC islet GAD cDNA clones.
 SQ Sequence 32 BP; 6 A; 16 C; 6 G; 4 T;

 Query Match 100.0%; Score 18; DB 1; Length 32;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Oy 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

 RESULT 11
 ID Q36668 standard; cDNA: 35 BP.
 AC Q36668;
 DT 09-JUN-1993 (first entry)
 DE PPO adapter primer #1.
 KW Polyphenol oxidase; PPO; catalyst; browning; fruit; plastid; vacuole;
 KW transform: coffee; tea; black olives; grapevine; chloroplast; apple;
 KW transit peptide; recombinant plasmid; PCR; primer; amplify; broad bean;
 KW potato; polymerase chain reaction; ss.
 OS Synthetic.
 PN WO9302195-A.
 PD 04-FEB-1993.
 PF 16-JUL-1992; AU0356.
 PR 17-JUL-1991; AU-007248.
 PA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PI Dry IB, Robinson SP;
 DR WPI: 93-058792/07.
 PT DNA encoding polyphenol oxidase polypeptide or fragment - useful
 PT for modifying the oxidase activity in fruit and vegetables to
 PT decrease or enhance browning
 PS Claim 15; Page 23; 44pp; English.
 CC The sequences given in Q36668-69 are adaptors which were used in the
 CC cloning of the polyphenol oxidase (PPO) enzyme genes from various
 CC plants. The PPO genes were isolated, and recombinant plasmids for
 CC transformation of plant cells were produced by PCR using the primers
 CC given in Q36678-78. PPO is thought to be the predominant catalyst in
 CC browning of fruit caused by injury or damage. PPO is localised in the
 CC plastids of plant cells whereas the phenolic substrates of the enzyme
 CC are stored in the plant cell vacuole. This compartmentation prevents
 CC the browning reaction from occurring unless the plant cells are damaged
 CC and the enzyme and the substrate are mixed. The PPO gene sequences
 CC could be used to construct synthetic genes which may be used to
 CC transform plants to decrease expression of the enzyme gene. In some
 CC instances, eg. coffee, tea, black olives etc., it is desirable to
 CC increase the level of PPO to produce desired levels of browning or
 CC changes in flavour compounds. The grapevine PPO gene codes for an
 CC additional 103 amino acids upstream of the N-terminus of the mature
 CC protein. This region has the properties of a chloroplast transit

CC peptide and is most likely responsible for targeting of the protein
 CC to be imported into the chloroplast and processed to produce mature
 CC PPO. Transformation of plants with this gene may therefore result
 CC in correct targeting and maturation of the grapevine PPO in other
 CC species and result in accumulation of active grapevine PPO enzyme in
 CC the plastids of these tissues.
 SQ Sequence 35 BP; 5 A; 5 C; 20 G; 5 T;

 Query Match 100.0%; Score 18; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Oy 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

 RESULT 12
 ID Q48416 standard; DNA: 18 BP.
 AC Q48416;
 DT 09-MAR-1994 (first entry)
 DE PCR primer for adenosine receptor partial cDNA (clone hval-3a).
 KW Adenosine receptor; expression system; cardiovascular disorders;
 KW renal disorders; neurological disorders; assay; ss.
 OS Synthetic.
 PN GB2264948-A.
 PD 15-SEP-1993.
 PF 05-MAR-1993; 004582.
 PR 13-MAR-1992; US-850701.
 PR 13-MAR-1992; US-850702.
 PR 13-MAR-1992; US-850707.
 PR 15-JAN-1993; US-005945.
 PA (MERI) MERCK & CO INC.
 DR WPI: 93-290442/37.
 PT Human adenosine receptor proteins - and expression systems,
 PT including cells for use in adenosine receptor binding assay
 PT Disclosure; Page 15; 65pp; English.
 CC Purified adenosine receptor proteins can be used in assays to screen
 CC for therapeutic human adenosine receptor agonists, antagonists or
 CC binding enhancers. These agonists, antagonists and binding
 CC enhancers may be used in the treatment of cardiovascular, renal and
 CC neurological disorders. Two copies of this primer were used to
 CC amplify partial cDNA of an adenosine receptor clone (hval-3a) for
 CC its subsequent insertion into an expression vector.
 SQ Sequence 18 BP; 5 A; 5 C; 5 G; 3 T;

 Query Match 100.0%; Score 18; DB 1; Length 18;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

 Oy 1 GACTCGAGTCGACATCGA 18
 ||||||||||||||||
 Db 1 GACTCGAGTCGACATCGA 18

 RESULT 13
 ID Q57547 standard; cDNA: 39 BP.
 AC Q57547;
 DT 20-MAY-1994 (first entry)
 DE Oligonucleotide ZC2487, an RNA -> cDNA primer.
 KW Gutamine:fructose-6-phosphate amidotransferase; insulin resistance;
 KW diabetes; screening; antagonists; GFAr; ss.
 OS Synthetic.
 PN WO9321330-A.
 PD 28-OCT-1993.
 PF 22-APR-1993; U03773.
 PR 22-APR-1992; US-872648.
 PR 05-NOV-1992; US-973330.
 PA (ZYMO) ZYMOGENETICS INC.

PI McKernan PA, McKnight GL, Ohara PJ, Sheppard PO;
 PI Smith RA, Vanness J;
 DR WPI: 93-351740/44.
 PT Human glutamine-fructose-6-phosphate amido transferase and
 PT corresp. DNA - for screening antagonists which inhibit insulin
 PT resistance, for use in diabetes patients
 PS Example 1: Page 61; 77pp; English.
 CC The sequence is a primer for the preparation of cDNA from RNA
 CC obtained from human pheochromocytoma cells. The cDNA was amplified by
 CC PCR to obtain the human glutamine:fructose-6-phosphate aminotransferase
 CC gene. The gene and its product are useful for screening for its
 CC antagonists which inhibit its exogenous activity. Such antagonists
 CC inhibit insulin resistance and are useful in the treatment of
 CC diabetes patients.
 CC See also Q57548-63 and Q51188.
 SQ Sequence 39 BP; 6 A; 6 C; 6 G; 21 T;

Query Match 100.0%; Score 18; DB 1; Length 39;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGACTCGACATCGA 18
 Db 1 GACTCGACTCGACATCGA 18
 |||||||

RESULT 14
 ID 052503 standard; DNA: 35 BP.
 AC Q52503;
 DT 31-MAY-1994 (first entry)
 DE Helminth aminopeptidase H11-1 adaptor primer.
 KW Helminth; aminopeptidase; H11-3; H11-2; H11-1; Integral membrane;
 KW alpha-amino acyl peptide hydrolase (microsome); multigene family;
 KW antigen; vaccine; parasite; human; virus; microbe; PCR; primer;
 KW polymerase chain reaction; amplify; Haemonchus contortus; ss.
 OS Synthetic.
 PN WO9323542-A.
 PD 25-NOV-1993.
 PF 07-MAY-1993; G00943.
 PR 08-MAY-1992; GB-009993.
 PA (AGRI-) AGRIC & FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ;
 PI Smith TS;
 DR WPI: 93-386574/48.
 PT New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 PS Example: Page 32; 137pp; English.
 CC The sequences given in Q52503-11 are primers which were used to
 CC amplify cDNA clones of the helminth aminopeptidase genes H11-1, -2
 CC and -3. The amplified sequences encode H110D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between the amplified clones can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H110D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in
 CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by the H11 genes may be used in vaccines to stimulate immune
 CC response against helminth parasites in humans or other animals. These
 CC DNA sequences may be incorporated into a virus or microbe and used in
 CC a similar manner.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 18; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGACTCGACATCGA 18
 Db 1 GACTCGACTCGACATCGA 18
 ..|||||

Db 1 GACTCGACTCGACATCGA 18
 ID 059015 standard; DNA: 35 BP.
 AC Q59015;
 DT 12-APR-1994 (first entry)
 DE RACE PCR adaptor-dr.
 KW Phosphonositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphonositides;
 KW ss; amplification; p110.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphonositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1: Page 39; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSVneo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dr as primers. Products were fractionated using an agarose gel. The
 CC DNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also Q51155-6, Q59012-23 and Q57522-3.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 100.0%; Score 18; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.33;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GACTCGACTCGACATCGA 18
 Db 1 GACTCGACTCGACATCGA 18
 |||||||

Search completed: September 13, 1999, 15:59:31
 Job time: 4798 sec

THIS PAGE BLANK (USP 12)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds

(without alignments)
20.952 Million cell updates/sec

Title: US-09-325-095-18

Perfect score: 18
Sequence: 1 GACTCGAGTCGACATCGA 18

Scoring table: IDENTITY_NUC

Searched: 2546578 segs, 986266752 residues

Database : 0

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	18	100.0	337	32	AA340759	AA340759 EST46034
2	15	83.3	360	24	D75133	D75133 CELK096P8F
3	15	83.3	242	32	AA331366	AA331366 EST35191
4	15	83.3	535	39	AA897864	AA897864 NCM7A9T3
5	15	83.3	822	51	A1723540	A1723540 hcg15IE3.
6	14.8	82.2	390	36	C72810	C72810 C72810 Rice
7	14.8	82.2	251	43	A1224703	A1224703 gw98h10.x
8	14.4	80.0	323	20	T18855	T18855 a03001r Tes
9	14.4	80.0	360	35	C50284	C50284 C50284 Yuj1
10	14.4	80.0	494	39	AA846538	AA846538 g156f12.s
11	14.4	80.0	246	49	AV015355	AV015355 AV015355
12	13.8	76.7	312	21	D48281	D48281 R1C514415A
13	13.8	76.7	377	25	N81968	N81968 TGE5TZY6190
14	13.8	76.7	357	26	W46073	W46073 mc79f07.r1
15	13.8	76.7	412	26	W97957	W97957 mg05d09.r1
16	13.8	76.7	260	27	AA049420	AA049420 mj34a06.r
17	13.8	76.7	432	29	AA189627	AA189627 mt90d09.r
18	13.8	76.7	332	30	AA198427	AA198427 mw46c01.r
19	13.8	76.7	335	30	AA237205	AA237205 mx18d03.r
20	13.8	76.7	425	30	AA238027	AA238027 mx16b10.r
21	13.8	76.7	402	30	AA245117	AA245117 mw28d02.r
22	13.8	76.7	447	30	AA245606	AA245606 mx01f01.r
23	13.8	76.7	451	30	AA253764	AA253764 mw03c07.r
24	13.8	76.7	426	30	AA254869	AA254869 mt84907.r
25	13.8	76.7	434	30	AA272892	AA272892 va39h02.r
26	13.8	76.7	396	33	AA403552	AA403552 vdt8d01.r
27	13.8	76.7	332	33	AA412997	AA412997 vdb6f09.r
28	13.8	76.7	440	34	AA470223	AA470223 vdb0f05.r
29	13.8	76.7	420	34	AA495288	AA495288 fa02907.r
30	13.8	76.7	388	36	AA600575	AA600575 vm75f06.r
31	13.8	76.7	178	36	AA637510	AA637510 vc32d10.r
32	13.8	76.7	313	37	AA709591	AA709591 vc37f03.r
33	13.8	76.7	401	38	AA739275	AA739275 vv51f11.r
34	13.8	76.7	297	39	AA835087	AA835087 vk63a09.s
35	13.8	76.7	317	39	AA869732	AA869732 vq15b08.r
36	13.8	76.7	286	40	AA966805	AA966805 sg011a1.f
37	13.8	76.7	421	41	A1048842	A1048842 uc84a04.Y
38	13.8	76.7	807	41	A1068903	A1068903 mrs00004C
39	13.8	76.7	366	42	A1130122	A1130122 SMOYLCAN
40	13.8	76.7	316	43	A1222065	A1222065 q102c11.x
41	13.8	76.7	293	44	A1286399	A1286399 u175y09.Y
42	13.8	76.7	482	46	A1399018	A1399018 NCM10A5T3
43	13.8	76.7	587	47	A1534610	A1534610 SD07520.5
44	13.8	76.7	426	48	A1545100	A1545100 fb70503.Y
45	13.4	74.4	540	51	A0070229	A0070229 A0070229

ALIGNMENTS

RESULT 1
AA340759
LOCUS
DEFINITION EST46034 Fetal kidney II Homo sapiens cDNA 5' end similar to
similar to protein phosphatase 2a, beta, 55 kDa, mRNA sequence.
ACCESSION AA340759
NTD 91992998
VERSION AA340759.1 GI:1992998

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 337)
Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,
Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geochagen N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Sauder D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferrite A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,
Raymond L., Wei Y.F., Wing J., Xu C., Yu G.L., Ruben S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,
Fraser C.M. and Venter J.C.
*Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl.), 3-174 (1995)
On Sep 12, 1996 this sequence version replaced gi:1406936.
Other ESTs: EST46035 THC125647
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
5712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
Seq primer: M13-21.
Location/Qualifiers
1..337
/organism="Homo sapiens"
/db_xref="ATCC (inhost):142407"
/db_xref="taxon:9606"
/clone_lib="Fetal Kidney II"
/dev_stage="fetus"
/note="Organ: Kidney; Vector: pBluescript KS-; Site_1:
XhoI; Site_2: EcoRI"
BASE COUNT 63 a 88 c 95 g 88 t 3 others
ORIGIN

Query Match 100.0%; Score 18; DB 32; Length 337;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACTCGAGTCGACATCGA 18
|||||
Db 4 GACTCGAGTCGACATCGA 21

RESULT 2
D75133 360 bp mRNA EST 14-DEC-1995
LOCUS CELK096F8F Yuj1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk96f8 5', mRNA sequence.
D75133
ACCESSION
D75133
NID 91120917
VERSION 91120917
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Rhoditina; Rhoditoidae; Rhoditidae; Pelodierinae; Caenorhabditis.
AUTHORS 1 (bases 1 to 360)
Kohara Y., Mitsuki H., Nishigaki A., Mochizuki T., Sugimoto A. and
Tabara H.
Toward an expression map of the C. elegans genome
Unpublished (1994)
On Sep 21, 1992 this sequence version replaced gi:276359.
TITLE
JOURNAL
COMMENT
Contact: Yuj1 Kohara
National Institute of Genetics
Gene Library Lab
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykoha@ddj.nig.ac.j
Insert length: 1628 Std Error: 0.00
High quality sequence stop: 345.
Location/Qualifiers
1..360
/organism="Caenorhabditis elegans"
/strain="CB1489 him-8(e1489)"
/db_xref="taxon:6239"
/clone_lib="Yuj1 Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue=whole animal"
BASE COUNT 77 a 103 c 99 g 79 t 2 others
ORIGIN

Query Match 83.3%; Score 15; DB 24; Length 360;
Best Local Similarity 93.8%; Pred. No. 1.5e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CTCGAGTCGACATCGA 18
|||||
Db 249 CTCGAGTCGACATCGA 234

RESULT 3
AA331366 242 bp mRNA EST 21-APR-1997
LOCUS EST35191 Embryo, 7 week, subtracted (total cDNA) II Homo sapiens
DEFINITION cDNA 5' end similar to similar to DNA polymerase III, mRNA
sequence.
AA331366
NID 91983609
VERSION AA331366.1 GI:1983609
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 242)
Adams M.D., Kerlavage A.R., Fleischmann R.D., Fuldner R.A.,
Bult C.J., Lee N.H., Kirkness E.F., Weinstock K.G., Gocayne J.D.,
White O., Sutton G., Blake J.A., Brandon R.C., Man-Wai C.,
Clayton R.A., Cline T.R., Cotton M.D., Earle-Hughes J., Fine L.D.,
Fitzgerald L.M., Fitzhugh W.M., Fritchman J.L., Geochagen N.S.,
Glodek A., Gnehm C.L., Hanna M.C., Hedblom E., Hinkle P.S., Jr.,
Kelley J.M., Kelley J.C., Liu L.-I., Marmaros S.M., Merrick J.M.,
Moreno-Palauques R.F., McDonald L.A., Nguyen D.T., Pelligrino S.M.,
Phillips C.A., Ryder S.E., Scott J.L., Sauder D.M., Shirley R.,
Small K.V., Spriggs T.A., Utterback T.R., Weidman J.F., Li Y.,
Bednarek D.P., Cao L., Cepeda M.A., Coleman T.A., Collins E.J.,
Dimke D., Feng D.-F., Ferrite A., Fischer C., Hastings G.A.,
He W.W., Hu J.S., Greene J.M., Gruber J., Hudson P., Kim A.K.,
Kozak D.L., Kunsch C., Hungjun J., Li H., Weissner P.S., Olsen H.,
Raymond L., Wei Y.F., Wing J., Xu C., Yu G.L., Ruben S.M.,
Dillon P.J., Fannon M.R., Rosen C.A., Haseltine W.A., Fields C.,
Fraser C.M. and Venter J.C.
*Initial assessment of human gene diversity and expression patterns

TITLE

JOURNAL
MEDLINE
COMMENT

based upon 83 million nucleotides of cDNA sequence
Nature 377 (5547 Suppl), 3-174 (1995)
On Sep 12, 1996 this sequence version replaced gi:1404942.

Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423

Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
Seq primer: M13 Reverse.

FEATURES

Location/Qualifiers

1..242
/organism="Homo sapiens"
/db_xref="ATCC (inhost):132998"
/db_xref="taxon:9606"
/clone_lib="Embryo, 7 week, subtracted (total cDNA) II"
/dev_stage="embryo, 7 wks"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI"

BASE COUNT 45 a 71 c 75 g 45 t 6 others
ORIGIN

Query Match 83.3%; Score 15; DB 32; Length 242;
Best Local Similarity 88.2%; Pred. No. 1.3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 2 ACTCGAGTCGACATCGA 18
|||||
Db 1 ACTNGAGTNGACATCGA 17

RESULT 4

AA897864 535 bp mRNA EST 12-APR-1998
LOCUS NCM7A973 Mycelial Neurospora crassa cDNA clone NCM7A9 5' end, mRNA
DEFINITION

ACCESSION AA897864

VERSION AA897864.1 GI:3044297

KEYWORDS EST.

SOURCE Neurospora crassa.

ORGANISM

Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
Sordariales; Sordariaceae; Neurospora.

REFERENCE

AUTHORS

1 (bases 1 to 535)
Leonard, P.M., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L.,
Nelson, M.A., Mitchell, J., Armljo, A.M., Bean, L., Bluyes, E.,
Cushing, T., Errett, A., Fleharty, M., Gorman, M., Judson, K.,
Miller, R., Ortega, J., Pavlova, I., Petea, J., Todisco, S.,
Trujillo, R., Valentine, J., Wells, A., Werner-Washburne, M., Yazzie, S.
and Natvig, D.O.
Expressed sequences from confidial, mycelial, and sexual stages of
Neurospora crassa

TITLE

Fungal Genet. Biol. 21, 348-363 (1997)

JOURNAL

97435349

On Jan 19, 1998 this sequence version replaced gi:2152245.

Contact: Natvig, D.O./Nelson, M.A.

Department of Biology

University of New Mexico

Cascenter Hall, Albuquerque, NM 87131, USA

Tel: 505 277 3411

Fax: 505 277 0304

Email: ngevbiology.unm.edu
Deposited in GSDb at the National Center for Genome Resources with
accession GSDb:S:116735

FEATURES
source
Seq primer: T3
Location/Qualifiers

1..535
/organism="Neurospora crassa"
/strain="74A"
/db_xref="taxon:5141"
/map="4p16.3; 2; X"
/clone_lib="NM7A9"
/clone_lib="Mycelial"
/sex="Mating type A"
/tissue_type="Mycelium"
/dev_stage="Mycelium"

BASE COUNT 136 a 176 c 109 g 114 t
ORIGIN

Query Match 83.3%; Score 15; DB 39; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACTCGAGTCGACATC 16
|||||
Db 40 ACTCGAGTCGACATC 54

RESULT 5

LOCUS A1723540 822 bp mRNA EST 10-JUN-1999
DEFINITION hcglis1E3.T7 Haemonchus contortus intestinal mRNA Haemonchus
contortus cDNA clone hcglis1E3.T7 T7 similar to 60S Ribosomal
protein L10, mRNA sequence.

ACCESSION A1723540

NTID 95041869

VERSION A1723540.1 GI:5041869

KEYWORDS EST.

SOURCE Haemonchus contortus.

ORGANISM Haemonchus contortus.

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Strongylida;

Trichostrongylidae; Trichostrongylidae; Haemonchidae; Haemonchus.

1 (bases 1 to 822)
Jasmer, D.P., Myler, P.J. and Roth, J.P.

Haemonchus contortus Intestinal EST Project
Unpublished (1999)

On Dec 20, 1995 this sequence version replaced gi:1133538.

Contact: Jasmer, D.P.
Department of Veterinary Microbiology and Pathology
Washington State University
Bustad 380, Pullman, WA 99164-7040, USA
Tel: 509-335-6040
Fax: 509-335-8528
Email: djasmer@vetmed.wsu.edu.

FEATURES

source

1..822
Location/Qualifiers

/organism="Haemonchus contortus"
/db_xref="taxon:6289"
/clone_lib="hcglis1E3.T7"
/clone_lib="Haemonchus contortus Intestinal mRNA"
/tissue_type="Intestine"

BASE COUNT 240 a 207 c 211 g 156 t 8 others
ORIGIN

Query Match 83.3%; Score 15; DB 51; Length 822;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CTCGAGTCGACATCG 17
|||||
Db 517 CTCGAGTCGACATCG 531

RESULT 6
LOCUS C72810 390 bp mRNA EST 22-SEP-1997
DEFINITION C72810 Rice panicle at flowering stage Oryza sativa cDNA clone
E2310.1A, mRNA sequence.
ACCESSION C72810
NID 92428347
VERSION C72810.1 GI:2428347
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
REFERENCE 1 (bases 1 to 390)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from panicle at flowering stage
JOURNAL Unpublished (1996)
COMMENT
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@agr.affrc.go.jp.
Location/Qualifiers
1.390
/organism="Oryza sativa"
/strain="Nippondare"
/db_xref="taxon:4530"
/clone_1lb="E2310.1A"
/dev_stage="Rice panicle at flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering
stage"
BASE COUNT 89 a 105 c 86 g 108 t 2 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 36; Length 390;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|| |||||
Db 283 GACTCGAGTCGACATCGA 300

RESULT 7
LOCUS A1224703 251 bp mRNA EST 21-DEC-1998
DEFINITION A1224703.1 NCI-CGAP_Brl4 Homo sapiens cDNA IMAGE:1999171 3',
similar to gb:L22154 60S RIBOSOMAL PROTEIN L37A (HUMAN);, mRNA
sequence.
ACCESSION A1224703
NID 93807416
VERSION A1224703.1 GI:3807416
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 251)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1995 this sequence version replaced gi:811445.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
unknown library type
Insert Length: 205 Std Error: 0.00
Seq primer: -40UP from G1bco.
Location/Qualifiers
1.251
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="NCI-CGAP_Brl4"
/sex="female"
/tissue_type="normal epithelium"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: breast; Vector: PAMPI; mRNA made from breast
epithelium, cDNA made by oligo-dT priming. Directionally
cloned. Size selected on agarose gel, average insert
size 500 bp. Primary library, non-amplified."
BASE COUNT 88 a 44 c 53 g 49 t 17 others
ORIGIN

Query Match 82.2%; Score 14.8; DB 43; Length 251;
Best Local Similarity 88.9%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCGA 18
|| |||||
Db 153 GATTCGCGTCGACATCGA 136

RESULT 8
LOCUS T18855 323 bp mRNA EST 25-SEP-1996
DEFINITION T18855 a03001r Testis 1 Homo sapiens cDNA clone a03001 3' end, mRNA
sequence.
ACCESSION T18855
NID 9600897
VERSION T18855.1 GI:600897
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 323)
AUTHORS Pawlak,A., Tousseint,C., Levy,I., Bulle,F., Poyard,M., Barouki,R.
and Gueliaen,G.
TITLE Characterization of a large population of mRNAs from human testis
JOURNAL Genomics 26, 151-158 (1995)
MEDLINE 95301283
COMMENT On Sep 21, 1992 this sequence version replaced gi:276479.
Contact: Gueliaen G
Unité INSERM 99
INSERM
Unité INSERM 99, Hôpital Henri Mondor, 94010 Creteil, France
Tel: (33)149813530
Fax: (33)148980908
Email: gueliaen@infoblogen.fr
Seq primer: M13 reverse.
Location/Qualifiers
1.323
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1lb="Testis 1"
/note="Vector: pGEM 5zf(+); Site 1: EcoRV; site 2: NotI;
mRNA was prepared from human testis of a 27 years old man.
cDNA was prepared using a 15mer oligo dT anchored by two
degenerated bases at its 3' end and containing a NotI site

at its 5' end. The cDNA was cloned between EcoRV and NotI sites of pGEM 5zf(+). The 3' end is at the NotI site. The EcoRV site is lost during the cloning procedure. cDNA corresponding to abundant species were eliminated from this library.

BASE COUNT 73 a 91 c 63 g 96 t

Query Match 80.0%; Score 14.4; DB 20; Length 323;
Best Local Similarity 93.8%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATC 16
|||||
Db 290 GACTCGAGTCGACATC 305

RESULT 9

LOCUS C50284 360 bp mRNA EST 11-SEP-1997
DEFINITION C50284 Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk479h2 5', mRNA sequence.

ACCESSION C50284
NID g2387537
VERSION C50284.1 GI:2387537
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

REFERENCE 1 (bases 1 to 360)
AUTHORS Kohara, Y., Mochizuki, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.
TITLE Expression map of the C. elegans genome
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394150.

FEATURES
source Contact: Yui Kohara
Gene Library Lab
National Institute of Genetics
Yata 111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.jp.
Location/Qualifiers

1. 360
/organism="Caenorhabditis elegans"
/strain="Cbl469 him-8(e1489)"
/db_xref="taxon:6239"
/clone="yk479h2"
/clone_lib="Yui Kohara unpublished cDNA"
/note="dev stage-varied, sex=Hermaphrodite male,
tissue type=whole animal"

BASE COUNT 77 a 104 c 97 g 82 t

Query Match 80.0%; Score 14.4; DB 35; Length 360;
Best Local Similarity 93.8%; Pred. No. 3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTCGAGTCGACATCGA 18
|||||
Db 261 CTCGAGTCGACATCGA 246

RESULT 10

LOCUS AA846538 494 bp mRNA EST 31-DEC-1998
DEFINITION a156f12.s1 Soares testis_NHT Homo sapiens cDNA clone IMAGE:1394351
3', mRNA sequence.
ACCESSION AA846538

NID g2932678
VERSION AA846538.1 GI:2932678
KEYWORDS EST.
SOURCE
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 494)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151248.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldi, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/ILM at:
www.bio.lnl.gov/db/brp/image/image.html

Insert Length: 592 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 483.

FEATURES

source Location/Qualifiers

1. 494
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21q"
/clone="IMAGE:1394351"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: p7773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTCGAGCGCCGCCCAATTTTGTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified p7773 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldi."

BASE COUNT 130 a 127 c 85 g 152 t

Query Match 80.0%; Score 14.4; DB 39; Length 494;
Best Local Similarity 93.8%; Pred. No. 3.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATC 16
|||||
Db 292 GACTCGAGTCGACATC 307

RESULT 11
LOCUS AV015355 246 bp mRNA EST 03-JUN-1999
DEFINITION AV015355 Mus musculus 18-day embryo C57BL/6J Mus musculus cDNA
clone 1110059C18, mRNA sequence.
ACCESSION AV015355
NID g4792347
VERSION AV015355.1 GI:4792347
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

SEQUENCE
AUTHORS

Eutheria: Rodentia: Sciurognathi; Muridae: Murinae; Mus.
1 (bases 1 to 246)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Itawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuna, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigenoto, Y., Shiraki, T., Sogabe, Y.,
Suganaga, Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,
Tomihaga, S., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
On Mar 20, 1998 this sequence version replaced gi:2980208.

TITLE
JOURNAL
COMMENT

Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-resetc.riken.go.jp
Thermotabilization and thermostabilization of thermostable enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.riken.go.jp) for
further details.

FEATURES
source

1..246
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="6"
/clone="1110059C18"
/clone_lib="Mus musculus 18-day embryo C57BL/6J"
/sex="mixed"
/dev_stage="18-day embryo"
BASE COUNT 65 a 56 c 58 g 67 t
ORIGIN

Query Match 80.0%; Score 14.4; DB 49; Length 246;
Best Local Similarity 93.8%; Pred. No. 2.8e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTCGAGTCGACATCGA 18
|||||
DB 29 CTCGAGTCGACATCGA 14

RESULT 12
D48281 312 bp mRNA EST 02-AUG-1995
LOCUS R1C514415A Rice green shoot Oryza sativa cDNA, mRNA sequence.
DEFINITION D48281
ACCESSION 5701990
NID D48281.1 GI:701990
VERSION
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
1 (bases 1 to 312)
Sasaki, T., Miyao, A. and Yamamoto, K.
Rice cDNA from callus 1995
Unpublished (1995)
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba

Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abrr.affrc.go.jp
High quality sequence stop: 297.
Location/Qualifiers
1..312
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="Rice green shoot"
/note="Green shoot (8 days old)"
BASE COUNT 53 a 99 c 88 g 71 t 1 others
ORIGIN

Query Match 76.7%; Score 13.8; DB 21; Length 312;
Best Local Similarity 88.2%; Pred. No. 6.1e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GACTCGAGTCGACATCG 17
|||||
DB 160 GACTCGAGTCGACGCG 176

RESULT 13
N81968/c 377 bp mRNA EST 10-SEP-1997
LOCUS T9ESTZy61g09.r1 TgrH Tachyzoite cDNA Toxoplasma gondii cDNA clone
DEFINITION t9zy61g09.r1 5', mRNA sequence.
ACCESSION N81968
NID 91257721
VERSION N81968.1 GI:1257721
KEYWORDS EST.
SOURCE Toxoplasma gondii.
ORGANISM Toxoplasma gondii.
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Toxoplasma.
1 (bases 1 to 377)
Hehl, A., Manger, I., Marra, M., Sidley, L.D., Ajlola, J.A.,
Aslett, M.A., Dietrich, N., Dubuque, T., Hillier, L., Kucaba, T.,
Wan, K.L., Waterston, R.H. and Boothroyd, J.
WashU-Merck-Stanford-NIH Toxoplasma EST project
Unpublished (1996)
On Sep 21, 1992 this sequence version replaced gi:276215.
Contact: Marra M
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: toxo@wustl.wustl.edu
Library can be obtained from Genome Systems Inc. (genomemo.net);
Seq primer: T3
High quality sequence stop: 297.
Location/Qualifiers
1..377
/organism="Toxoplasma gondii"
/strain="RH"
/db_xref="taxon:5811"
/clone="t9zy61g09.r1"
/clone_lib="TgrH Tachyzoite cDNA"
/lap_host="XLI-Blue MRF"
/note="Vector: lambda ZAP; site_1: EcoRI; site_2: XhoI;
Toxoplasma RH strain tachyzoites were grown in human
foreskin fibroblast cultures in vitro. The library was
constructed by K.L. Wan, Cambridge University. cDNAs were
synthesized from polyA RNAs by oligo d(T) priming and
directionally cloned into the EcoRI to XhoI sites of the
lambda zapII vector using the ZAP-cDNA synthesis kit.

FEATURES
source

(Stratagene). WARNING: the library contains a small percentage of cDNAs derived from the human host cells.

BASE COUNT 102 a 78 c 111 g 84 t 2 others

ORIGIN

Query Match 76.7% Score 13.8; DB 25; Length 377;
Best Local Similarity 88.2%; Pred. No. 6.4e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ACTGAGTCGACATCGA 18
|||||
Db 340 ACTGCGTCGACATCGA 324

RESULT 14
W46073 357 bp mRNA EST 23-MAY-1996
LOCUS m079f07.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:354757 5' similar to PIR:S28237 S28237 NADH
dehydrogenase ;, mRNA sequence.
W46073
ACCESSION 91330807
NID W46073.1 GI:1330807
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 357)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Oct 18, 1995 this sequence version replaced gi:1023145.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Oct 18, 1995 this sequence version replaced gi:1023145.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through ILNLT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:226557
Putative full length read
Seq primer: ETPRimer
High quality sequence stop: 343.
Location/Qualifiers

FEATURES

source
1. 357
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:354757"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGCGCGGAAATTTTATTTTATTTTATTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo.

BASE COUNT 89 a 98 c 91 g 79 t

ORIGIN

Query Match 76.7% Score 13.8; DB 26; Length 357;
Best Local Similarity 88.2%; Pred. No. 6.3e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 ACTGAGTCGACATCGA 18
|||||
Db 177 ACGCGTCGACATCGA 193

RESULT 15
W97957 412 bp mRNA EST 16-JUL-1996
LOCUS m05d09.r1 Soares mouse embryo NBME13.5 14.5 Mus musculus cDNA
DEFINITION clone IMAGE:422897 5' similar to PIR:S28237 S28237 NADH
dehydrogenase ;, mRNA sequence.
W97957
ACCESSION 91427866
NID W97957.1 GI:1427866
VERSION EST.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 412)
Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On May 9, 1995 this sequence version replaced gi:803042.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 9, 1995 this sequence version replaced gi:803042.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through ILNLT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:257449
Seq primer: mob.REGA+ET
High quality sequence stop: 329.
Location/Qualifiers

FEATURES

source
1. 412
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="19"
/clone="IMAGE:422897"
/clone_lib="Soares mouse embryo NBME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGCGCGGAAATTTTATTTTATTTTATTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of

normalization, and was constructed by Bento Soares and
M. Fatima Bonaldo.
BASE COUNT 111 a 108 c 100 g 93 t
ORIGIN

Query Match 76.7%; Score 13.8; DB 26; Length 412;
Best Local Similarity 88.2%; Pred. No. 6.5e+02;
Matches 15; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 2 ACTCGAGTCGACATCGA 18
||| |||||
Db 167 ACCCGTGTGACATCGA 183

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec


```

/codon_start=1
/protein_id="CA02284.1"
/db_xref="PID:6304808"
/db_xref="PID:q2294346"
/db_xref="GI:2294346"
/translation="MPRRSGEINGIHPRIIVCELLPNCMTVTECLREALTIT
IKHELFKARKKYPHLOLDESSYIVSTQENREFEDEFERLDDLFPPLKVI
EVGVNKEKILNREGEFALGMVCEPDVMDKDEVDERNINLVCKEAVDLNDSPE
SRAMVYPPNVESSPELPHIYNKDKQIIVYVIVSPNNDKQYTLKINHDSPE
IYAEIRKRTKSMLESEQLKCYEYQKILKCGDEFELEKPSYQKIRISC
IYAGRPNTMLMAKESYLSQDPMDCFTMSYSRISSTAPVNGESTSLVYNSAL
RKLICATVYVNIIDIKIYRTGIVHGEPLCNVYVNPQPCSPRNENLNDIY
ITDLPRAATLCISYKGRKQKKECHCEPLANGINLFTPTLTIVSGKALMLPPEH
GLEDLNPITGVTSNPKETPCLELEDFSSVYKFPDVSVEEASNSVSEAGSY
SHAGLSNRLARDNELENDKEDQRAICTRDPLSEITQEKDLMSHRHCVTIPILP
KLLSYKNSRDVQMYCIVADMPPIKPEQAMLELDCYPPPMVGFAYRCLKELT
DKLSYLIQVQVLYKEQYLDNLVRLKALTNORIGHEFMHLSKEMKATVSO
REGILSYCGRACGWYKILNVOVEMELINUTDILQEKDEKQKQKRIYVOMR
RDEPDALOGFISPLPAHQNLNREBRINSASRPMINWENPDIHSELHPONE
IIFKNGDDLQOMLTLQIIRIMENIWOQGLDRLMPLYCGLSIGCVGLIEVNSHT
IMQIOCKGLKALQFNSTHLQWLKDNKKGFIYDAIDLFTRSCAGCVATFIIGIG
DHNSNIMWKDGOLFHDGFHDKKKKFFYKREYVFFVLTODFLIYISKAGECT
KTRERFERQECYKATLAIROHANLEINFSMVLGSGMPELQSFODIAYIRYTLADLK
TQEPALFYPMKMDAHGSGMTXMDMIFHTIKAHAN"

```

```

BASE COUNT      1147 a      626 c      722 g      1003 t
ORIGIN
Query Match      100.0% Score 20; DB 5; Length 3498;
Best Local Similarity 100.0% Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 TGCTGTAATTCCTAATGCTG 20
          |||||
Db      3325 TCCTGTAATTCCTAATGCTG 3306

```

```

RESULT 2
LOCUS      AC004707.c      DNA      PRI      23-JUN-1998
DEFINITION Homo sapiens chromosome 17, clone hRPC.117_B.12, complete sequence.
ACCESSION      AC004707
NID      63249127
VERSION      AC004707.1 GI:3249127
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens

```

```

REFERENCE      Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
                Eutheria; Primates; Catarrhini; Hominoidea; Homo.
                1 (bases 1 to 123585)
                Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.
                Homo sapiens chromosome 17, clone hRPC.117_B.12
                Unpublished
                2 (bases 1 to 123585)

```

```

REFERENCE      Birren,B., Baldwin,J., Barne,N., Beckerly,R., Benn,J., Boatin,C.,
                Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
                Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
                Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
                Gensheimer,S., Geraldery,K., Gilmartin,T., Grant,G., Hagos,B.,
                Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L.,
                Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
                Melgrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
                Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
                Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
                Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
                Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
                Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
                Zody,M.
                Direct Submission
                Submitted (20-MAY-1998) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
                3 (bases 1 to 123585)

```

```

AUTHORS      Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,
                Baker,J., Baldwin,J., Barne,N., Beckerly,R., Benn,J., Boatin,C.,
                Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,
                Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,
                Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,
                Gensheimer,S., Geraldery,K., Gilmartin,T., Grant,G., Hagos,B.,
                Harris,K., Horton,L., Howland,J.C., Hul,L., Jacotot,L., Kann,L.,
                Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
                Melgrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,
                Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B.,
                Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,
                Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
                Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,
                Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and
                Zody,M.
                Direct Submission
                Submitted (23-JUN-1998) Whitehead Institute/MIT Center for Genome
                Research, 320 Charles Street, Cambridge, MA 02141, USA
                On Jun 23, 1998 this sequence version replaced gi:3237292.
                All repeats were identified using RepeatMasker: Smit, A.F.A. &
                Green, P. (1996-1997)
                http://ftp.genome.washington.edu/RM/RepeatMasker.html.
                Location/Qualifiers
                source      1..123585
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="hRPC.117_B.12"
                /clone_11b="Peter de Jong/ human PAC library"
                /map="17"
                /chromosome="17"
                complement(8..159)
                /rpt_family="MIR"
                repeat_region      216..479
                /rpt_family="AluDb"
                repeat_region      1591..1853
                /rpt_family="L1P3"
                complement(1874..1995)
                /rpt_family="FLAM_A"
                repeat_region      2317..2615
                /rpt_family="AluDo"
                complement(2756..2816)
                /rpt_family="MER2"
                repeat_region      2824..3127
                /rpt_family="AluSx"
                repeat_region      3207..3264
                /rpt_family="L1MC/D"
                complement(3459..3624)
                /rpt_family="MER2"
                complement(4022..4322)
                /rpt_family="AluDo"
                repeat_region      4738..4758
                /rpt_family="AT-rich"
                repeat_region      4959..5259
                /rpt_family="AluSp"
                repeat_region      5262..5562
                /rpt_family="AluY"
                repeat_region      5567..5595
                /rpt_family="(CAA)n"
                repeat_region      5774..5898
                /rpt_family="FLAM_A"
                complement(5928..5965)
                /rpt_family="AT-rich"
                repeat_region      5974..6280
                /rpt_family="AluDb"
                repeat_region      6283..6406
                /rpt_family="L1MB3"
                repeat_region      6521..6822
                /rpt_family="AluSp"
                repeat_region      6825..6862
                /rpt_family="L1MB3"
                complement(6995..7285)
                /rpt_family="AluSg"
                complement(7354..7657)

```



```

repeat_region /rpt_family="AluIo"
767..8039
/rpt_family="AluIb"
8176..8237
/Note="Single-stranded terminator coverage."
repeat_region 8237..8305
/rpt_family="FRAM/FAM"
8307..8584
/rpt_family="AluSg"
complement(8556..8942)
/rpt_family="MIR"
9488..9789
/rpt_family="AluIo"
9900..10138
/rpt_family="AluIo"
complement(11010..11304)
/rpt_family="AluSg"
11873..12174
/rpt_family="AluSp"
complement(12490..12583)
/rpt_family="MIR"
complement(12890..13187)
/rpt_family="AluSx"
complement(13192..13485)
/rpt_family="AluSx"
complement(13486..13647)
/rpt_family="FRAM"
complement(13972..14011)
/rpt_family="MER57B"
15475..15514
/rpt_family="GC_rich"
15354..15588
/rpt_family="GC_rich"
complement(17197..17234)
/rpt_family="MIR"
17255..17376
/rpt_family="MIR"
complement(18275..18303)
/rpt_family="CATAn"
18304..18402
/rpt_family="AluIo/FRAM"
18409..19117
/rpt_family="LMB3"
19192..19492
/rpt_family="AluY"
19530..19562
/rpt_family="AT_rich"
complement(20163..20471)
/rpt_family="AluIo"
complement(21003..21135)
/rpt_family="MLTID"
21136..21222
/rpt_family="AluSg"
21223..21684
/rpt_family="LTR2"
21685..22570
/rpt_family="HERVE"
22576..23322
/rpt_family="HERV17"
23287..23359
/rpt_family="MER57_internal"
23359..23650
/rpt_family="HERVKC4"
23675..24109
/rpt_family="MER57_internal"
24240..25000
/rpt_family="BAEV_int"
24999..25779
/rpt_family="HERV1"
25776..26352
/rpt_family="HERVE"
26349..28539
/rpt_family="HERVE"

```

```

repeat_region 28540..28990
/rpt_family="LTR2"
29009..29207
/rpt_family="AluSg"
complement(29211..29359)
/rpt_family="MLTID"
complement(29535..29714)
/rpt_family="L2"
29812..29907
/rpt_family="AluY"
29908..30133
/rpt_family="AluIb"
30139..30438
/rpt_family="AluSx"
complement(30563..30626)
/rpt_family="(TAA)n"
complement(30734..30796)
/rpt_family="(TA)n"
complement(30802..30921)
/rpt_family="FLAM_A"

Query Match 87.0% Score 17.4; DB 11; Length 123585;
Best Local Similarity 94.7% Pred. No. 78;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCCTGAATTCATATGCT 19
Db 95937 TGCCTGAATTCATATGCT 95919

RESULT 3
AF126284 11824 bp RNA VRL 09-FEB-1999
LOCUS AF126284
DEFINITION Aura virus polypeptide 1 and polypeptide 2 genes, complete cds.
ACCESSION AF126284 S78478
NID 94240567
VERSION AF126284.1 GI:4240567
KEYWORDS
SOURCE
Aura virus.
Aura virus.
ORGANISM
Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
Alphavirus.
REFERENCE
1 (bases 1 to 11824)
Rumenapf,T., Strauss,E.G. and Strauss,J.H.
TITLE
Aura virus is a New World representative of Sindbis-like viruses
JOURNAL Virology 208 (2), 621-633 (1995)
MEDLINE 95266268
REFERENCE
2 (bases 1 to 11824)
Rumenapf,T.H.
TITLE
Direct Submission
JOURNAL Submitted (08-FEB-1999) Vet. Virology, Justus-Liebig University,
Frankfurter Str. 107, Giessen D-35392, Germany
On Feb 9, 1999 this sequence version replaced gi:999057.

COMMENT
FEATURES
source
1..11824
/organism="Aura virus"
/db_xref="taxon:44158"
join(78..5723,5727..7577)
/Note="contains leaky opal termination codon"
/codon_start=1
/product="polyprotein 1"
/protein_id="AADI3622.1"
/db_xref="PID:94240568"
/db_xref="GI:4240568"
/translation="MERPTVHVDDVQSPVLOKSPQFEIYAQOVTNDAHARA
FSHASKLIEHETSTYIIDIGSAPRRATSEKRYCQPMRSPEDPRMAYASL
ADKAGETNKLHDLKLDLSVLESPDAETGTCFHDVICTRTAIVSVMQNTINP
STYHQLKGVRLYWGFDPTQGFSSMAGSYSTYTNMADEVLARNGICSTLR
RECTMGRLSTFRKRLKPGTNVFSVSTIYPENRADLOSWHLPVHLGKOSTLR
CDIAPNCEGVYVKRTITSPGJTGVRNVTYNNSEGLCKITDTPKGERVSPVCTY
IPSPICOMGTILATIDOPDAOKLVLGNRIYVNNKTNNNTMTQNYLLPAYAGTL
SKAKERKACSDSEKPLNVRERKLAFGLIAFKKTIHSPRPFGTGTIYKVAEFA
FPMSSVTTISLPSLRKVKLLVTKTKNPVTTITDAVKAQAEATNEAVEIAEEK"

```

AAAPPLKPTAPVADVCETDLDVDDGAALVPRPKKIIIPQEGCVRGSTVI
SPRAVLNNOLEPIHELAQVAKIITHGGTGRISVPIPAKALLPGCPMSQHPAL
SSAQLVYNEREFNLKRHLIAITGAUKATPEBOYKVCADIDHEVYVDARCKV
RBAHQGLVYVGLINPPIHELAEGITAPAPYHETIGTPEGSKALIKSTVT
LMDLVTSGKKNCKEIVNDOKKRGMTINTRTVDLLGNKAKVAVLVDENAFCHA
GTLMLAIVKPRKRVIGDPRKMPFIMOLKVNENPBDICSTHYKISRPT
QVPTALVSTLHDGKRTTNPCKRAIEIDVNGSTRKRDYILTCRGWYKOGIDYF
GGGHDNRASQGLTRGVAVAKOVNENLIEKSHVAVLLTRDRLVMTLQDDE
WIKYLLNKGNETALIEMOQHEDIMAINSTSVDFEASKVTCWAKAIIPLR
TAGIELTPEQMEDLPQFRNDQPYSVYALDVICTMFGMDSSGIFSPETLPEP
AVGRVRAHMDSPGCGKFGYKNAVPTCKKVPYIRAKGQOILPITGRVSPARN
NIVPLNRNLPHSITAGLKKAAPLKHPINOLPGHMLIVSKETCCVSKRTWAPL
CYRGADHNDLHRGPPLSRDYLVYNNKOPYRHHOCEENHAGMLRTIASALNCL
KKGITLAKAIGFADNSDVALSLAKRVASAVRSTQFTEMEFVRKQDNDRE
ROFTOHNLAVSNIDNTKDSGAAPSVYRVMANADTEAVVANAARAKONGV
CAIEFKMKPSEENATTEVEYAVMKCHKRVYHAGPDFRPTLEAKRLONAH
VAKIVKNGISSVAIPILSTGIAGADLDLSLRCILFALDRIDADVITICDKKE
OHADAIRMEBOYTELKDPIEDIEGLTVHDSCLKHIGYSTQYGLKXVEGTRKE
HOTADIAIRALPPOVANOICITYTGEPMESREKCPEDSPASAPRTIPLC
MAAMTERICRVRNSVITTCSSPLKRIKNOQIQCTKVLVFNDDVPIIPAR
VYINDEPVTPTDSDPPTCSSRLTPTLNAESDVLSTFSEIDSLSLNEPAR
HVMISFKLRATYAIQALPQKLSMRREDTTPROPVPYPRKRAALSLANOLNLR
HATISSVOAEVYHNGSTPEALNERSILKRPVPLRKOITNLSRLANOLNMP
ITEGDEAGEDELRLTSPPTPEGDSOEMOEFENRQYILRSDGQYH
OOKSVONSTTEILLERSLEKIHAPVULQKEMKCYOHSPTYANSQSKRVE
NKATYTGGLDGLKATVPDVABECYKTRIPMTSASVPREFSPEVAVCNNE
HENYPTVASIYQITDEYDALDWNESGSLDITATPCPALRSEFTHSLPTLRSAV
PSAFONTLONVLSAATKNCNONTOMRELVLDSAVENVCFFKYACNDYMEFEKP
IRITTECVSYVARALPQKLEALAFKTHQVLOEYPMOFWMDKADYKVPGRHT
ERRKVOVIOAAPELATAYLIGIHRELVRITAVLLPNTHTLDSAEDEFDITANF
SYVHPLETIDISFKSDSDSLATLMLTLEDJGUDRLTLEAFSGEITVHPLTA
TFPRGAMKSKGHELTLYNTVYLVNIVASRVLEORLROSKCAFTGDNDITGVSDK
INADRCATWNNVEKIIDAVIGIKAPYEGGGLLEOVVHTACRSDPKLRIKFKGP
LFDVDEQDHRRLALEDTRAMFRVGIQELKKAESRYEVOEQVPLALATFERSD
KAFKALRSGSRHLVYGGPK"

7628. 11362
/codon_start=1
/product="polypeptide 2"
/protein_id="A01362.1"
/db_xref="PID:g4240569"
/db_xref="GI:4240569"
/translation="MNSVFNPFGRGAVYOPPIAMPRRRAAPAPPSGLTQIOOLT
RAVRALVILNDRORPARTRPRKTKOKPKPKONOKPOOQKNGKNOPOPKPK
PKRORLAKFPADRTFVKNEDKMGAVAMEKONVIRPLVKTGIDTIDPALAKLFT
KSSSTDMERAKLPTKMSDAEITTEHPEVFTNMHGAVOFGSRRITLPTGVGPGDS
GRLIDNSGRVAVIYVGNANVEPTALSVYTNKGAALIKTHEDIVEMSRATMCI
LONVTPDCRPTCYNRNBDLTLMLETNNHSDYVLDALRCPTRRHVSSTPTDD
FTLAPYGLICHRCKTMEPCYSPRIKIEKWDADQVLRIOVSAOLGYRATASAR
LRFMGAGVPEIIOEGALIDEKVTSPKCLHSHKGFVYVRCPPGDSITTSILKAGSD
OCTIIPMYGVKFGVREKTYLTPMHSTQIPCLTIEPRKSKAGYTMHAPGOSITML
MEESGEVAVOPTSGANVTYECGDFKGTIVTARKIDGCTERQCLASDHAKW
FNSPLIRHTDQAGKLAIPFLDOAOCTVPLAPGVKAYRMSLSLASHDHTLL
LTRHNGENPOPAEMVYVSVTNFSTIGFEYTMGNOKPRVYAOESAPHPCHPM
EARNIHVPYTOAPSGFEKWNNGOPLDTPAGFGCVYVNRADCAVSLPIVDI
PAAATRVSEPLPSLICKTVSTCTDGTAVGVYVNSDRAGOCVAVHSHSTAVLD
PSVYEOKGETTLKSTRSLQADFEVSKGCTITTCACQCPTEHVMNRPNQSTPDS
SHKSTSMWITLALMGGISIAIYVIALVLTFAHR"

BASE COUNT 3462 a 2906 c 2828 g 2628 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 11824;
Best Local Similarity 94.7%; Pred. No. 1.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 TCGTGAATCTAATGCT 19
|||||

Db 11350 TCGTGAATCTAATGCT 11332

RESULT 4
FLA3N8HAH/c RNA VRL 23-OCT-1998
LOCUS Influenza A virus gene for hemagglutinin precursor (hemagglutinin 1 region), partial cds, isolate: A/Equine/Rome/5/91.
DEFINITION
ACCESSION D30684
MID 9487670
VERSION 3.0
KEYWORDS hemagglutinin precursor; HA1; hemagglutinin 1.
SOURCE Influenza A virus (strain:H3N8, isolate:A/Equine/Rome/5/91) cDNA to genomic RNA.
ORGANISM Influenza A virus
VIRUSES: ssRNA negative-strand viruses; Orthomyxoviridae; Influenza virus A and B group; Influenza A viruses.

REFERENCE 1 (bases 1 to 1090)
Lindstrom, S.E., Endo, A., Pecoraro, M.R., Sugita, S., Damiani, A., Himoto, Y., Kamata, M., Kumanomido, T., and Nerome, K.
Complete nucleotide sequence of the HA1 region of the hemagglutinin gene of A/Equine/Rome/5/91 (H3N8) influenza virus
Unpublished (1994)
Submitted (10-May-1994) to DDBJ by:
Stephen E. Lindstrom
Virology I
National Institute of Health
1-23-1 Toyama, Shinjuku-ku
Tokyo 162
Japan
Phone: 03-5285-1111 x2531
Fax: 03-5285-1155.
LOCATION/Qualifiers
1. 1090
/organism="Influenza A virus"
/isolate="A/Equine/Rome/5/91"
/strain="H3N8"
/db_xref="taxon:11320"
/chromosome="segment 4"
10. 57
10. >1090
/codon_start=1
/product="hemagglutinin precursor"
/protein_id="BA33945.1"
/db_xref="PID:d1034920"
/db_xref="PID:g3779054"
/db_xref="GI:3779054"
/translation="MKTIILILTHWYSONPTSGNNTATLIGHHAVANGTLVTKI
TDQDEVNNAELVQSIGIKCNRSYVLDGRCTLDLMDGPRCHDDPRYEMWDL
IERSSANCIPTIDIPDASLRSIVASSGLTAEFTWGTGONSGSKCKGSD
SFFSLNLTGSGNSYPLNVTNNKKNFELYTWGIHPSNSNEOTKLYIOESGRVY
VSTERSOQTVIPNIGSRPLVNGOGGRISYWTIVKPGDILMNSNGNLVAPRGFKLR
TGKSSMGSDLIDTCVSECITTPNGSIPNDKPFQNVKVTYKCPKTIROWTLKATG
MRNPEKRIIFGALINGFINMGW"

FEATURES
source
sig_peptide
cds
mat_peptide
BASE COUNT 391 a 217 c 232 g 250 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 17; Length 1090;
Best Local Similarity 94.7%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 TCGTGAATCTAATGCT 19
|||||

Db 420 TCGTGAATCTAATGCT 402

RESULT 5
HS105D16 148984 bp DNA PRI 30-NOV-1998
LOCUS Human DNA sequence from clone 105D16 on chromosome Xp11.3-11.4
DEFINITION

Contains pseudogene similar to laminin-binding protein, CA repeat,
STS, complete sequence.

ACCESSION AL031311
NID 93947674
VERSION AL031311.1 GI:3947674
KEYWORDS HTG;
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 148984)
AUTHORS Pearce, A.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Dec 2, 1998 this sequence version replaced gi:3647153.

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence is the entire insert of clone 105D16. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/ChrX>
105D16 is from the library RPC16 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see <http://bacpac.med.buffalo.edu/VECTOR:RPC4>.

FEATURES

Source

1. 148984
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="105D16"
/map="p11-3-11.4"
/clone_11b="RPC16"
1. 249
/note="AluSp repeat: matches 59. .312 of consensus"
858. .905
/note="4 copies 12 mer 83% conserved"
1004. .1128
/note="FLAM_C repeat: matches 1. .125 of consensus"
1617. .1914
/note="AluSx repeat: matches 1. .299 of consensus"
2129. 2293
/note="L1McC repeat: matches 1253. .1420 of consensus"
2535. .2839
/note="AluSx repeat: matches 1. .305 of consensus"
2844. .3155
/note="AluSx repeat: matches 1. .311 of consensus"
3161. 3454
/note="HAL1 repeat: matches 547. .841 of consensus"
3207. .3472
/note="L1McC repeat: matches 1669. .1937 of consensus"
3490. .4104
/note="L1M4 repeat: matches 2370. .3002 of consensus"
4105. 4287
/note="MIR1A2 repeat: matches 8. .201 of consensus"
4341. .4641
/note="AluY repeat: matches 1. .298 of consensus"
5048. .5177
/note="L1M4 repeat: matches 5666. .5809 of consensus"
5285. 5387
/note="L1M4 repeat: matches 5434. .5530 of consensus"

repeat_region 5395. .5691
/note="AluSx repeat: matches 1. .297 of consensus"
repeat_region 5724. .6028
/note="AluSc repeat: matches 1. .308 of consensus"
repeat_region 6038. .6086
/note="AluSc repeat: matches 247. .295 of consensus"
6106. .6249
/note="L1M4 repeat: matches 5269. .5413 of consensus"
6363. .6629
/note="L1McS repeat: matches 7691. .7925 of consensus"
6876. .7193
/note="L1M4 repeat: matches 4477. .4785 of consensus"
7194. .7504
/note="AluSx repeat: matches 1. .311 of consensus"
7505. .7762
/note="L1M4 repeat: matches 4185. .4477 of consensus"
7763. .8097
/note="AluSg repeat: matches 1. .310 of consensus"
8098. .8205
/note="L1M4 repeat: matches 4081. .4185 of consensus"
8559. .8775
/note="HERV1 repeat: matches 5247. .5474 of consensus"
9055. .9080
/note="13 copies 2 mer ac 92% conserved"
9094. .9386
/note="AluSp repeat: matches 1. .294 of consensus"
9762. .10119
/note="THE1B repeat: matches 1. .382 of consensus"
10380. .10678
/note="AluSg repeat: matches 7. .306 of consensus"
10712. .11232
/note="L1Mc2 repeat: matches 5620. .6155 of consensus"
11218. .11621
/note="L1M4 repeat: matches 4955. .5351 of consensus"
11622. .11906
/note="AluSx repeat: matches 1. .286 of consensus"
11907. .12497
/note="L1M4 repeat: matches 4339. .4955 of consensus"
12494. .12543
/note="L1M4 repeat: matches 3102. .3154 of consensus"
12544. .12853
/note="AluSg repeat: matches 1. .306 of consensus"
12854. .13043
/note="L1M4 repeat: matches 2933. .3102 of consensus"
13070. .13224
/note="MIR1B repeat: matches 210. .390 of consensus"
14565. .14588
/note="12 copies 2 mer gt 96% conserved"
15584. .15895
/note="L1MB8 repeat: matches 5507. .5860 of consensus"
15913. .16075
/note="L1MB8 repeat: matches 6001. .6173 of consensus"
16092. .16394
/note="AluS8 repeat: matches 1. .302 of consensus"
16594. .16882
/note="AluSg repeat: matches 1. .296 of consensus"
17546. .17703
/note="MIR repeat: matches 61. .226 of consensus"
17736. .17993
/note="L1PA3 repeat: matches 5889. .6148 of consensus"
18033. .18340
/note="AluSx repeat: matches 1. .310 of consensus"
18509. .18757
/note="AluSg repeat: matches 6. .271 of consensus"
19384. .19412
/note="MIR repeat: matches 236. .262 of consensus"
19413. .19624
/note="THE1B repeat: matches 171. .364 of consensus"
19625. .19928
/note="AluSx repeat: matches 1. .304 of consensus"
19929. .20091
/note="THE1B repeat: matches 3. .171 of consensus"
20092. .20260
repeat_region


```

repeat_region /note="MIR repeat: matches 30. .261 of consensus"
21765. .22188
/note="L2 repeat: matches 1687. .2104 of consensus"
repeat_region 22189. .22339
/note="BURI repeat: matches 1801. .1952 of consensus"
22575. .22880
/note="AlusX repeat: matches 1. .306 of consensus"
25611. .25779
/note="MER5B repeat: matches 1. .171 of consensus"
27119. .27239
/note="L1M5 repeat: matches 6180. .6300 of consensus"
27355. .27628
/note="L2 repeat: matches 2458. .2749 of consensus"
27657. .27963
/note="AluIo repeat: matches 10. .312 of consensus"
28121. .28508
/note="L1M10 repeat: matches 5979. .6319 of consensus"
28549. .29091
/note="L1M1 repeat: matches 5624. .6165 of consensus"
29100. .29189
/note="45 copies 2 mer ta 71% conserved"
29240. .29628
/note="L1M1 repeat: matches 5146. .5403 of consensus"
31238. .31401
/note="FRAM repeat: matches 1. .162 of consensus"
32068. .32179
/note="MIR repeat: matches 125. .235 of consensus"
32372. .32615
/note="MIR repeat: matches 1. .252 of consensus"
32853. .33009
/note="MER20 repeat: matches 1. .161 of consensus"
33187. .33283
/note="L2 repeat: matches 2584. .2700 of consensus"
33685. .33913
/note="MIR repeat: matches 31. .260 of consensus"
33918. .34091
/note="MIR repeat: matches 86. .260 of consensus"
34475. .34830
/note="MLR1C repeat: matches 3. .318 of consensus"
35067. .35214
/note="L2 repeat: matches 2585. .2723 of consensus"
35461. .35754
/note="AlusG repeat: matches 1. .294 of consensus"
35830. .35959
/note="MIR repeat: matches 9. .142 of consensus"
36023. .36229
/note="L2 repeat: matches 2076. .2290 of consensus"
36348. .36709
/note="MLR1D repeat: matches 1. .398 of consensus"
38052. .38255
/note="L1M9 repeat: matches 6101. .6308 of consensus"
38257. .39342
/note="TIGER1 repeat: matches 2. .1164 of consensus"
39343. .41236
/note="L1P2 repeat: matches 4227. .6146 of consensus"
41252. .42476
/note="TIGER1 repeat: matches 1166. .2418 of consensus"
42479. .44326
/note="L1M7 repeat: matches 4254. .6100 of consensus"
44314. .44903
/note="L1M4C repeat: matches 1467. .2414 of consensus"
44915. .45072
/note="L1M3 repeat: matches 5919. .6090 of consensus"
45071. .45764
/note="L1M3 repeat: matches 1153. .1785 of consensus"
45765. .46081
/note="AluIb8 repeat: matches 1. .318 of consensus"
46082. .46160
/note="L1M3 repeat: matches 1082. .1153 of consensus"
46161. .46587
/note="MSTa repeat: matches 1. .426 of consensus"
46588. .48395
/note="L1M3 repeat: matches -879. .1082 of consensus"

```

```

repeat_region 48598. .48800
/note="L2 repeat: matches 2541. .2750 of consensus"
48951. .49292
/note="L2 repeat: matches 1143. .1537 of consensus"
49553. .49661
/note="L2 repeat: matches 2089. .2195 of consensus"
49811. .49834
/note="L2 repeat: matches 2089. .2195 of consensus"
49837. .49905
/note="L2 repeat: matches 52. .117 of consensus"
49908. .49976
/note="MER5B repeat: matches 39. .108 of consensus"
50021. .50078
/note="L2 repeat: matches 39. .108 of consensus"
50079. .50113
/note="MER5B repeat: matches 2386. .2420 of consensus"
50501. .50743
/note="MIR repeat: matches 3. .262 of consensus"
54352. .54535
/note="L2 repeat: matches 2317. .2508 of consensus"
54858. .54934
/note="MLR1H repeat: matches 448. .525 of consensus"
55056. .55285
/note="MLR1H repeat: matches 4. .279 of consensus"
56812. .56820
/note="MER81 repeat: matches 1. .114 of consensus"
57445. .57502
/note="29 copies 2 mer tt 74% conserved"
57963. .58321
/note="L2 repeat: matches 2414. .2749 of consensus"
58856. .58943
/note="MIR repeat: matches 48. .139 of consensus"
59154. .59687
/note="L1R12 repeat: matches 149. .671 of consensus"
59714. .60013
/note="AlusX repeat: matches 3. .302 of consensus"
60674. .61024
/note="THERC repeat: matches 1. .371 of consensus"
61221. .61728
/note="MLR1D repeat: matches 6. .505 of consensus"
62466. .62941
/note="HERV1 repeat: matches 3873. .4366 of consensus"
63839. .63878
/note="20 copies 2 mer gt 100% conserved"
65738. .70935

```

```

Query Match 84.0% Score 16.8; DB 10; Length 70935;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 TGCTGTAATTCATATGCTG 20
DB 30626 TGATGTATATCTATGCTG 30607

```

```

RESULT 7
AC004468/C AC004468 127757 bp DNA PRI 10-NOV-1998
LOCUS Homo sapiens Xp21 pAC RPII-37A12 containing exons 10 to 16 of the
DEFINITION Duchenne Muscular Dystrophy gene, complete sequence.
ACCESSION AC004468
NID 93097815
VERSION AC004468.1 GI:3097815
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 127757)
Muzny,D., Arenson,A.D., Adams,C., Bunac,C., Carvelli,K., Chang,J.,
Chacko,J., Chen,J., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
Ganesh,R., Garcia,C., Goodman,M., Gotrell,J.H., Haywood,M.,
Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpaty,S., Kovar,C.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 127757)
Submitted (21-MAR-1998) Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
Worley, K.C.
Direct Submission
Submitted (01-MAY-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 127757)
Worley, K.C.
Direct Submission
Submitted (12-MAY-1998) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 127757)
Gorrell, L.L.
Direct Submission
Submitted (13-OCT-1998) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
6 (bases 1 to 127757)
Gorrell, L.L.
Direct Submission
Submitted (03-NOV-1998) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
7 (bases 1 to 127757)
Gorrell, L.L.
Direct Submission
Submitted (06-NOV-1998) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
8 (bases 1 to 127757)
Gorrell, L.L.
Direct Submission
Submitted (10-NOV-1998) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, U.S.A.
On May 1, 1998 this sequence version replaced gi:3063478.
Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui Zhang.
Exon/intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.
FEATURES
SOURCE
Location/Qualifiers
1..127757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPCL1-37A12"
/chromosome="X"
/clone_id="Roswell1 Park Cancer Center Human PAC library"
598..663
/rpt_family="(CA)n"
692..757
/rpt_family="MIR"
complement(882..1384)

```

```

/note="Insertion of C."
/clone="Insertion with respect to Y13187."
misc_difference 35340
/note="Deletion of G."
/clone="Deletion with respect to Y13187."
misc_difference 35358
/note="Deletion of A."
/clone="Deletion with respect to Y13187."
misc_difference 35386
/note="Deletion of A."
/clone="Deletion with respect to Y13187."
misc_difference 35413
/note="Deletion of G."
/clone="Deletion with respect to Y13187."
misc_difference 36366
/note="Deletion of T."
/clone="Deletion with respect to Y13187."
misc_difference 36893
/note="Deletion of CT."
/clone="Deletion with respect to Y13187."
misc_difference 36906
/note="Deletion of CT."
/clone="Deletion with respect to Y13187."
misc_difference 36907
/note="Deletion of CT."
/clone="Deletion with respect to Y13187."
misc_difference 36980
/note="Deletion of C."
/clone="Deletion with respect to Y13187."
repeat_region /rpt_family="MAD1"
complement(37591..37670)
repeat_region /rpt_family="TAAn"
complement(37685..37779)
repeat_region /rpt_family="L2"
complement(37810..38338)
repeat_region /rpt_family="L2"
complement(38367..38448)
repeat_region /rpt_family="L2"
complement(38891..39104)
repeat_region /rpt_family="AT rich"
complement(39105..39230)
repeat_region /rpt_family="FLAM_C"

Query Match      84.0%; Score 16.8; DB 11; Length 127757;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGTCTAATCTAATGCTG 20
|||||
Db 17931 TGCCTTAATTAATGCTG 17912

RESULT 8
AC004674 158886 bp DNA PRI 18-NOV-1998
LOCUS Homo sapiens Xp22 BAC GSHB-600G8 (Genome Systems Human BAC library)
DEFINITION complete sequence.
ACCESSION AC004674
NID 93273377
VERSION AC004674.1 GI:3273377
KEYWORDS HTG.
SOURCE Homo sapiens
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158886)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 158886)
REFERENCE
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 30, 1998 this sequence version replaced gi:3764539.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Location/Qualifiers
1..158886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GSHB-600G8"
/chromosome="X"
/clone_id="Genome Systems Human BAC library"
/map="Xp22"
complement(174..242)
/rpt_family="MER5B"
1904..1927
/rpt_family="CAAAA)n"
2946..3064
/rpt_family="(GGA)n"
6186..6364
/rpt_family="MER58A"
6380..6548
/rpt_family="MER5B"
6820..6982
/rpt_family="FRAM"
complement(8974..9232)
/rpt_family="LIME1"
9233..9529
/rpt_family="AluSx"
complement(9530..9807)
/rpt_family="LIME7"
complement(9817..9870)
/rpt_family="LIME1"
complement(9895..10264)
/rpt_family="LTR16C"
complement(10459..10770)
/rpt_family="Aluub"
12524..12641
/rpt_family="MER86"
12695..12749
/rpt_family="LIME"
14814..15117
/rpt_family="AluSx"
complement(15126..15240)
/rpt_family="FAM"

```

```

REFERENCE
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 158886)
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (30-JUN-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 158886)
REFERENCE
AUTHORS Morley,K.C.
TITLE Direct Submission
JOURNAL Submitted (18-NOV-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 30, 1998 this sequence version replaced gi:3764539.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Location/Qualifiers
1..158886
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GSHB-600G8"
/chromosome="X"
/clone_id="Genome Systems Human BAC library"
/map="Xp22"
complement(174..242)
/rpt_family="MER5B"
1904..1927
/rpt_family="CAAAA)n"
2946..3064
/rpt_family="(GGA)n"
6186..6364
/rpt_family="MER58A"
6380..6548
/rpt_family="MER5B"
6820..6982
/rpt_family="FRAM"
complement(8974..9232)
/rpt_family="LIME1"
9233..9529
/rpt_family="AluSx"
complement(9530..9807)
/rpt_family="LIME7"
complement(9817..9870)
/rpt_family="LIME1"
complement(9895..10264)
/rpt_family="LTR16C"
complement(10459..10770)
/rpt_family="Aluub"
12524..12641
/rpt_family="MER86"
12695..12749
/rpt_family="LIME"
14814..15117
/rpt_family="AluSx"
complement(15126..15240)
/rpt_family="FAM"

```

```

repeat_region 15590..15888
repeat_region /rpt_family="AluSp"
STS 15942..16947
/rpt_family="L1ME2"
16256..16405
standard_name="WI-10833, Chr. X, Homo sapiens"
/db_xref="dbSTS:16691"
repeat_region 17043..17224
/rpt_family="L1ME2"
repeat_region complement(17562..17860)
/rpt_family="Aluub"
repeat_region complement(18976..19270)
/rpt_family="AluSx"
repeat_region complement(19536..19589)
/rpt_family="(TA)n"
repeat_region 20280..20562
/rpt_family="Aluuo"
repeat_region 22503..22802
/rpt_family="Aluy"
repeat_region 23221..23448
/rpt_family="MIR"
repeat_region complement(23594..23645)
/rpt_family="L1MA10"
repeat_region complement(23637..23967)
/rpt_family="L1ME1"
repeat_region 23980..24049
/rpt_family="L2"
repeat_region 24383..24692
/rpt_family="Aluub"
repeat_region 25219..25342
/rpt_family="MIR"
repeat_region 25351..25780
/rpt_family="L1R16B"
repeat_region complement(25789..26097)
/rpt_family="Aluuo"
repeat_region 26156..26215
/rpt_family="L2"
repeat_region complement(26319..26587)
/rpt_family="L1MB4"
repeat_region complement(27108..27516)
/rpt_family="L1"
repeat_region 27520..27660
/rpt_family="L1MB2"
repeat_region 27662..27953
/rpt_family="AluSx"
repeat_region 27963..28071
/rpt_family="L1MB2"
repeat_region 28072..28105
/rpt_family="MER4D"
repeat_region 28108..28162
/rpt_family="L1"
repeat_region 28163..28273
/rpt_family="Aluub"
repeat_region 28313..28367
/rpt_family="L1P5"
repeat_region 28368..28474
/rpt_family="(CAT)n"
repeat_region 28475..28774
/rpt_family="L1PB2"
repeat_region 28944..29091
/rpt_family="MER4D"
repeat_region complement(29192..29314)
/rpt_family="MER34"
repeat_region 29338..29825
/rpt_family="L1MA9"
repeat_region complement(30099..30327)
/rpt_family="MER20"
gene join(30351..30445,32077..32250,34228..34281,35716..35797,
36473..36517,36631..36696)
/ene="AA412296
/ene="AA412296 zu10N02.s1 Soares_testis_NHT Homo sapiens
cDNA clone 731475 3'"
repeat_region 30715..31015

```

```

repeat_region /rpt_family="AluSx"
31832..32162
/rpt_family="MER61"
repeat_region complement(33303..33344)
/rpt_family="AT-rich"
repeat_region 33579..33688
/rpt_family="L2"
repeat_region 34905..35089
/rpt_family="Aluy"
repeat_region complement(36258..36422)
/rpt_family="FRAM"
repeat_region 37106..37402
/rpt_family="AluSx"
repeat_region complement(39936..40010)
/rpt_family="L1P3"
repeat_region 40004..40559
/rpt_family="L1P7"
repeat_region 40578..40612
/rpt_family="AT-rich"
repeat_region 40613..40634
/rpt_family="(GA)n"
repeat_region complement(40674..40969)
/rpt_family="Aluy"
misc_feature 51970..52439
/ene="Region: AA402429-2168e05.r1 Soares_testis_NHT Homo
sapiens cDNA clone 727520 5' similar to TR:G833928 G833928
MOD1 PROTEIN"
52323..52863
/ene="Region: Similar to AA746358 oa56f05.r1
NCI CGAP GCBI Homo sapiens cDNA clone IMAGE:1308993 5'"
53107..53141
repeat_region

misc_feature
misc_feature
misc_feature

Query Match 84.0%; Score 16.8; DB 11; Length 158886;
Best Local Similarity 90.0%; Pred. NO. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCGTGAATTCATGCGC 20
Db 3750 TCGTGAATTCATGATG 3769

RESULT 9
AC005939 179485 bp DNA PRI 14-NOV-1998
LOCUS Homo sapiens chromosome 17, clone hnpk.467_K17, complete sequence.
DEFINITION AC005939
ACCESSION G3873180
NID AC005939.1 GI:3873180
VERSION AC005939.1 GI:3873180
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 179485)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hnpk.467_K17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179485)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckler,R., Benn,J., Boultwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., Corliss,D., Depierre,E., Devon,K.,
Dodelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geradigery,K., Grant,G., Hagos,B., Heathford,A.,
Hareana,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatias,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrum,J., Molla,M., Morris,M., Morrow,J.,
Mychalackij,J., Nahr,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Supramanian,A., Tesfaye,S., Tichovolsky,N., Torrella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

```


TITLE Direct Submission
JOURNAL Submitted (05-NOV-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 179485)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Boutwell, C., Brown, A., Castle, A., Cerny, J., Colangelo, M., Collins, S., Collins, A., Cooke, P., Corliss, D., Depayre, E., Devon, K., Dewar, K., Donnell, P., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Gardyna, S., Geradgery, K., Grant, G., Hagos, B., Heathford, A., Herena, L., Horton, L., Howland, J. C., Jacotot, L., Jones, C., Kann, L., Karates, A., Lehoczy, J., Macdonald, P., Margulis, N., McEwan, P., McGurt, A., McKernan, K., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Nant, R., Naylor, J., Niloff, W., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stillwell, J., Stojanovic, N., Stone, C., Subramaniam, A., Tesfaye, S., Tichovolsky, N., Toriella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, Y., Wyman, D., Ye, W. J., Zhao, J. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (14-NOV-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Nov 14, 1998 this sequence version replaced g1:3868740.
 All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.

FEATURES
SOURCE Location/Qualifiers
 1..179485
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="hbrp.467_K17"
 /clone_id="RPCT-11 human BAC library"
 /map="17"
 /chromosome="17"
 complement(917..1000)
 repeat_region
 /rpt_family="L2"
 1100..1240
 /rpt_family="MER20"
 1404..1497
 /rpt_family="MER20"
 complement(1789..1988)
 repeat_region
 /rpt_family="MIR"
 2070..2227
 /rpt_family="MIR"
 complement(2651..3041)
 /rpt_family="MSTB"
 complement(3128..3159)
 repeat_region
 /rpt_family="(CA)n"
 3943..3964
 /rpt_family="AT-rich"
 4133..4175
 /rpt_family="(CAAA)n"
 complement(4821..5004)
 repeat_region
 /rpt_family="MERSA"
 5017..6384
 /rpt_family="L1PA15"
 6413..6692
 /rpt_family="AluSc"
 6880..7033
 /rpt_family="L2"
 7034..7829
 /rpt_family="L2"
 7830..8129
 /rpt_family="L2"
 8183..8424
 /rpt_family="MIR"
 9028..9062
 /rpt_family="AT-rich"
 complement(9592..9891)
 repeat_region
 /rpt_family="AluSc"
 10114..10209
 /rpt_family="MIR"
 complement(10551..10901)
 repeat_region

repeat_region
 /rpt_family="L1ME3A"
 complement(11268..11491)
 repeat_region
 /rpt_family="L1M4C"
 complement(13198..13412)
 repeat_region
 /rpt_family="L1MA8"
 complement(13451..13697)
 /rpt_family="HERV16"
 13745..14036
 /rpt_family="AluYb8"
 complement(14738..15649)
 repeat_region
 /rpt_family="Char11el"
 15652..15948
 /rpt_family="AluY"
 complement(17202..17408)
 repeat_region
 /rpt_family="AluSg/x"
 17653..17679
 /rpt_family="AT-rich"
 complement(17685..17979)
 repeat_region
 /rpt_family="AluSc"
 18686..18802
 /rpt_family="MIR"
 19410..19544
 /rpt_family="L2"
 20448..21217
 /rpt_family="L2"
 complement(22467..22610)
 repeat_region
 /rpt_family="MIR"
 complement(22839..22956)
 repeat_region
 /rpt_family="L2"
 complement(23628..23675)
 /rpt_family="MLT2"
 complement(23724..24012)
 /rpt_family="MLT2"
 complement(24022..24387)
 /rpt_family="THE1B"
 complement(24388..24566)
 /rpt_family="MLT2"
 24676..25147
 /rpt_family="LTR33"
 complement(25261..25486)
 /rpt_family="AluSg/x"
 complement(25504..25936)
 /rpt_family="LTR8"
 complement(26042..26297)
 /rpt_family="LTR8"
 complement(26308..26453)
 /rpt_family="MIR"
 26484..26524
 /rpt_family="AT-rich"
 27458..27478
 /rpt_family="AT-rich"
 complement(27880..27921)
 /rpt_family="purine-rich"
 28488..28528
 /rpt_family="AT-rich"
 28660..29194
 /rpt_family="L1PA15"
 29195..29234
 /rpt_family="(TAAA)n"
 29589..29621
 /rpt_family="AT-rich"
 30085..30106
 /rpt_family="AT-rich"
 30553..30592
 /rpt_family="AT-rich"
 32459..32634
 /rpt_family="AluY"
 32642..32948
 /rpt_family="AluY"
 34130..34183
 /rpt_family="AT-rich"
 complement(34345..34382)
 /rpt_family="(CAT)n"
 repeat_region

```

repeat_region complement(34755..35134)
repeat_region /rpt_family="MER57B"
repeat_region complement(35140..35725)
repeat_region /rpt_family="MER57-internal"
repeat_region complement(35993..36153)
repeat_region /rpt_family="HERV1"
repeat_region 36165..36199
repeat_region /rpt_family="(CAATT)n"
repeat_region complement(36200..36476)
repeat_region /rpt_family="AluJo"
repeat_region complement(36910..37139)
repeat_region /rpt_family="HERV16"
repeat_region 37162..37309
repeat_region /rpt_family="(TA)n"
repeat_region complement(37402..37687)
repeat_region /rpt_family="AluSp"
repeat_region 38963..38994
repeat_region /rpt_family="(GAAA)n"
repeat_region complement(39055..39216)
repeat_region /rpt_family="L2"
repeat_region 39848..39938
repeat_region /rpt_family="A-rich"
repeat_region 39939..40222
repeat_region /rpt_family="AluX"

Query Match 84.0%: Score 16.8; DB 11; Length 179485;
Best Local Similarity 90.0%: Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 TGCTGTAATTCATGCTG 20
    ||||| ||||| ||||| |||||
Db 162750 TGCTGAAATGCTAATGCTG 162769

```

```

RESULT 10
LOCUS CELY73C8C 36277 bp DNA INV 05-MAR-1999
DEFINITION Caenorhabditis elegans cosmid Y73C8C.
ACCESSION AF101318
VERSION 93808329
KEYWORDS AF101318.1 GI:3808329
SOURCE
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 36277)
Wilson, R., Almscough, R., Anderson, K., Baynes, C., Berks, M.,
Boulton, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, D., Laister, N.,
Lattelle, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Ricken, L., Roopa, A.,
Saunders, D., Showkeen, R., Smalton, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson, S., Pratt, J., and Wohlmann, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 36277)
Bemis, G., Courtney, L., and Wohlmann, P.
TITLE The sequence of C. elegans cosmid Y73C8C
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 36277)
TITLE Waterston, R.
JOURNAL Direct Submission
AUTHORS Submitted (26-Oct-1998) Department of Genetics, Washington
REFERENCE University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
AUTHORS 4 (bases 1 to 36277)
Waterston, R.

```

TITLE JOURNAL COMMENT

Direct Submission
Submitted (05-MAR-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: rwenematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

This is a portion of YAC Y73C8, sequenced to span the gap between
C50H11 and F02C9. The 5' clone is C50H11, 1561 bp overlap; 3' clone
is F02C9, 4400 bp overlap. Actual start of YAC Y73C8 is at base
position 37818 of CELC50H11; actual end is at 41552 of CELT28A11

NOTES:

Coding sequences below are predicted from computer analysis, using
the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES SOURCE

```

1..36277
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/clone="Y73C8C"
/chromosome="v"
/complement(2834..4971)
/gene="Y73C8C.4"
/complement(join(2834..3145,3402..3668,3755..4102,
4511..4684,4735..4971))
/gene="Y73C8C.4"
/codon_start=1
/evidence=not experimental
/protein_id="AAC69342.1"
/db_xref="PID:93808331"
/translation="MRLSICWVILFTCFNYIESKSTKLNSLNVCTOSEKAEINAK
TPYNNMLNENSIQVTEIKNTGKIPTEGCGPNCNEADKREYRKE
KIETTYPMDCVSKITAYLRGYPGSGFPLTRKDPDSRKRYSEBEVKIAR
TKDRALEIYDTERVDIMSTRKPHSHSDFHLVSRCPVADKMSHEPAPRLRAP
KSSLELTTHETTEAWRTFKRWSKPSVDEHLGKPLSPVQNASHEPAPRLRAP
KSCSEVRVYIMKRPTRSRFLAALYMMRGEGTAATAEAGBADLMDGADLKFGNMH
FOFPADFCMKSSYFSEDRNTEVEICNVYEDVFNCLNEVTRREDAKTFFCFIDE
KFHPDONTTELKKECKEIOVYMGCGDGLITLNDIMDNYQKNTSEGLL"
/complement(5950..8732)
/gene="Y73C8C.5"
/complement(join(5950..6008,6067..6515,6569..6944,
7624..7750,7803..8178,8242..8732))
/gene="Y73C8C.5"
/note="contains similarity to C. elegans olfactory
receptor ODR-10-related proteins"
/codon_start=1
/evidence=not experimental
/protein_id="AAC69341.1"
/db_xref="PID:93808330"
/db_xref="GI:3808330"

```

/translation="MSALTELESHVETVEFEASTENILITIFLAFVYKNGK
RMVYFSTLGLVFALETTAKPSPHNTNGSLITSTSELPETPSLMSGSAITII
VAFIAVOFVITLCLINMATKFDGKNTVWIIIPVPGAFVITISMFCAAEYS
SVYATISKNGLEVSKLPRYSMTASDGLRNVTIYLIGISISFNFITVC
YKMHISMKIELAKFSVKNQOLHROFKAIIYOIIGITPIPIEFALPIELK
IDVMOGNTLEALNIKOSLANSFGSINXIMIEKCFREIRREGGEGNSYSAPY
KRMIVFAAGLMSGEPASIKPTPHNNGTLFSAANPOLINISATETVSCFY
LITTSFISVLEFITYRILCLFDSKTRPFEGGIGMLPLPGICYSTIITLAPND
YTDYVDSVAENGLVNSELARFSLPYNSDSVIRHSIIFLLIASLINFHSII
FCSLKMFMNMRKELEKFSVONQLOROFIALVQALVMTFLITPAVPIIAELAP
IVGIEVDWQSGNLSLVGFYPPCDICISVALVSEYRNINKKISMCCKNNAOVSVINS
L"
complement(9050.10372)
/gene="Y7308C.6"
complement(join(9050.9126,9174.9622,9885.10372))
/gene="Y7308C.6"
/note="contains similarity to C. elegans olfactory
receptor ODR-10-related proteins"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69346.1"
/db_xref="PID:93808335"
/db_xref="GI:3808335"
/translation="MILYDLAGVFTKIOFCATFLINSRLITLITCYIREVGA
IYFALGMITSGLEVTAPRYTHNKASLSEFVYTNLPRVITCLMSMGCVIY
SFIVQVYRILCLFKAOKIKKFGVKSIFLITPMLGCVITYAMNLCADYAD
YRGLVSEERSMEISNLPREFMLPYADGSRMSLVPLCIPFLIGIHI111YCC
RMHSMKELAKFSTONKLOROFHALVDTGLTPIELVLPASIDILVPLAPFGR
DIDOTGNLFAFVGFPFDSIAFMIIYKEVYVKKHVLGCLGTQNVSAVEAF
VQS"
complement(10614.12486)
/gene="Y7308C.7"
complement(join(10614.10877,10978.11266,11315.12486))
/gene="Y7308C.7"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69347.1"
/db_xref="PID:93808336"
/db_xref="GI:3808336"
/translation="MCHLELVKMOEDMDVSLKEVLA
DKRIEVEYIEVTAKEFATPIPTTGGGCVIADPEYVYATNMTAKFERMD
SOFLIEEFNEFDYDTLRTGVTFLDKVDLENEKETLIGLITCEPTLRK
GKGFSTSENPANELYRLKLDIFPDICGOSIYFEKIKAKSKLSTGAGCLHOL
INCLAKIPKISETHSQLACSDPTIGCCICSDAQNPRSNSEPEVKISKDVSH
AELESMETKLELSRSNEKETOEPENDPKLSKRSKRSQDTAKPEKSNAC
PKCERASEYNIKLEDLLEKIEYKHLKSLHAQOSESSEKSOIAEKNEIEDIAQ
LQAOOIYVAKHNGENSRIPEGEKIPRSASETEVQILHEMLEOKLLEKPNPSKI
TTMSKLEIVSONRENTLIARELIIYFEROIRGYTKAEYEPFEKAPLILCAOLIKP
EPNEDLNECLVLEDMKEHEHTLKCLDKRKHYTGCAQKMPKVRKILPCIDAGLD
DNEPPALS"
complement(13005.16676)
/gene="Y7308C.8"
complement(join(13005.13704,13754.13837,13890.13976,
14030.15210,15770.16051,16111.16347,16416.16676))
/gene="Y7308C.8"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69348.1"
/db_xref="PID:93808337"
/db_xref="GI:3808337"
/translation="MFLNPTFFITSPSSLKLYCNTPBELPLPMDCTOKNPKYNI
QNLIDSSNGKLAVTGSALETESLNFREFPGSHKEFEENNAENDPAVQGMKRPYI
YKHDFVVIQVIAICVSPLRRETEVALATIASYKIRLPLTLEHVPFCOTSTAR
QOGRMNFNRIKIGESLQKNEYTEFEEMQKISKFEFMSVHDSLOVLRNFAK
KIPRDNELVYRMRAMSKIVLEVEKIEHONOMEYPRFAENENYVYVIRIFDKP
NKLIMAEHLVYVMOEHMDVSSFEKEVLEMKIATFPGQVAAVRSPPDEKNEVIT
PITAKYATPIPTITPGSLCILASDFINLIVNATKIKFOTKIDSPQFLIREFN
FIDFSGAKGVMEVDLNDVEYINKMFAFKTILITSCBPKAVNNTKTFREDLT
NELYKLDLIESEICQSDYFDMTRAKMLKLPNVEFGAIVHLOIDICRSTEDT
IPAFIHNOLASRIFGYRCPVADANTKADVAPPEELVEVEISPAVNEPETS
ESKRVSKKOKKORSENPINAKPEKOSGNACPRASEYNIKLDDRLKIEKHL
RKSILSQOSESSEKOKTAGNEIEHKAQVETTKRELLDOKSLEINKEKPTRIOM
EQILDHONALVKKSIVIERDEBETOLLAQOIAOONALIAQOKISIRNLEENKLMOK

Query Match 84.0% Score 16.8; DB 37; Length 36277;
Best Local Similarity 90.0% Pred. No. 1.7e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 1 TGCTGAATTCATATGCG 20
Db 8541 TCGGTAATTCATATGCTG 8560
RESULT 11
AE001173/C
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITILE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
FEATURES
SOURCE
gene
CDS
/translation="MFLNPTFFITSPSSLKLYCNTPBELPLPMDCTOKNPKYNI
QNLIDSSNGKLAVTGSALETESLNFREFPGSHKEFEENNAENDPAVQGMKRPYI
YKHDFVVIQVIAICVSPLRRETEVALATIASYKIRLPLTLEHVPFCOTSTAR
QOGRMNFNRIKIGESLQKNEYTEFEEMQKISKFEFMSVHDSLOVLRNFAK
KIPRDNELVYRMRAMSKIVLEVEKIEHONOMEYPRFAENENYVYVIRIFDKP
NKLIMAEHLVYVMOEHMDVSSFEKEVLEMKIATFPGQVAAVRSPPDEKNEVIT
PITAKYATPIPTITPGSLCILASDFINLIVNATKIKFOTKIDSPQFLIREFN
FIDFSGAKGVMEVDLNDVEYINKMFAFKTILITSCBPKAVNNTKTFREDLT
NELYKLDLIESEICQSDYFDMTRAKMLKLPNVEFGAIVHLOIDICRSTEDT
IPAFIHNOLASRIFGYRCPVADANTKADVAPPEELVEVEISPAVNEPETS
ESKRVSKKOKKORSENPINAKPEKOSGNACPRASEYNIKLDDRLKIEKHL
RKSILSQOSESSEKOKTAGNEIEHKAQVETTKRELLDOKSLEINKEKPTRIOM
EQILDHONALVKKSIVIERDEBETOLLAQOIAOONALIAQOKISIRNLEENKLMOK
KPLVATITQRTFEDTEKODILFENLKAQVLEVENPMKINKTKSKLOVKNRE
NKITSKEEFTTFFERQINGTYREVENRKITIRSNQISSDOIPELRNPTSDERLAKY
KDFEENAPLICAOGLTTPESQKSDLEDECVLDMMEHEHTLCSNCKROYHFG
CAQWFEVAKRCPFCNSGLDDDEFPVLY"
18125.20950
/gene="Y7308C.3"
join(18125.18394,18475.18720,18777.19058,19119.20161,
20210.20263,20314.20397,20449.20950)
/gene="Y7308C.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC69345.1"
/db_xref="PID:93808334"
/db_xref="GI:3808334"
/translation="MFLNPTFFITSPSSLKLYCNTPBELPLPMDCADKNEKLVN
QNLIGSNKIKMYGSAEELDSINTORFSGHELLGFELAEYSSPOLKMSKS
TYVYKHDFVILQGMIPYSPRARETEVAMATVATYIKIRSVLKSCEHVPVLAOSDL

identified by sequence similarity; putative"

codon_start=1
transl_table=11
product="glucose-6-phosphate isomerase (pgi)"
protein_id="AAC67084.1"
db_xref="pid:92688672"
/db_xref="gi:2688672"
translation="MFLMYKNLNELENFKILGIAPELVNTALYGRINKEYDITIEGDSVHVAHNSKQINETHHLKIPONLSDSEAMLIEKYEVLGEEKINISNRKYLHLHTROQIKADVIEENKEMREFPOSELEKINYFAKQJHSGIKSNKCRKFNVAQVIGSSISGIPRAKSYIKNAKYNALMLMGVPISTINDPESEPVLSIVDETLFVYQSSNTLETKANQQLINKLKLNGIKETKKQWITTLKQSMALIEKGLIEFPMHDSIGCSFESTSAVGILITLCTEKVANEKILGAVEADKRSKLNKNVKNOMASLALISITERNLNASTNCIIAAYKAMENFYLHQOLEMESNGSVARFNETINRYKTVRIWGGIGDVQHSFEOMLHGDTIVPMDFIGFNETQLEKEDVSDNSNDKAKANLAIQIIAFSGKESNKNKNFOERSALYSKELPELPAVIGALITSHYENKVMFEGFLINISFDEGVQGLKIANOLIKNDNRKDEVIYESYKRIIKRF"
complement(1684. .2085)
/gene="BB0731"
complement(1684. .2085)
/note="similar to GP:2351763 percent identity: 31.82; identified by sequence similarity; putative"
codon_start=1
transl_table=11
product="conserved hypothetical protein"
protein_id="AAC67083.1"
db_xref="pid:92688671"
/db_xref="gi:2688671"
translation="MLDIDQIKIKRRKRNIGDIEFLKNSIYHGLIPIITDKNNKLIAGNRRIQLEIGIKETIEKVAIVSTENKKLTLEIEDENNVRNSTRSENGEAYLKIYSSNNIIRFLKILIKIKMCKIRNRKI"
complement(2092. .4890)
/gene="BB0732"
complement(2092. .4890)
/note="similar to GP:1914832 percent identity: 29.81; identified by sequence similarity; putative"
codon_start=1
transl_table=11
product="penicillin-binding protein (pbp-3)"
protein_id="AAC67082.1"
db_xref="pid:92688670"
/db_xref="gi:2688670"
translation="MKLNSPMIKINETHKLIIVLYFAVSPSITLSSAVSKTINIOKDKNFVAVPNSRLIDNGIKOITQFISDENREMLPRKMPDNILITLLREDIGFSSHRGSLCIFRAPIYIVIGRYFSGSGSTIQDLALITLQNARSLIKLKEIHYAIQLEKSLIILEYILNKVYFGNGNTGIVAAESFEFGKSVKNINTISEVMITQLPNALEYSPLVNPEFSKIQAVLNOVNSGIVIAAEKFEWQNTDWTMAQTSALSNKKDQAPFESYIRQKILKYLPDGANIYKDQYSYSTLDELQAKYADKVNIMINKRATMNLNRSESIININSEIVPVADISDLGKIKNLRINGROYKRLKRPYEDNIDILASGALITGDKIDKATKEKIIKKNLTKPLIQAPBEAMAIQPTSAIRAMVSGSHRDNFENRATQAKQPSARFALFYFAAIDLKRTTAATMSDSFVLKNGEYVAPGNTGCMKGNVLRQALASLNTIPALRILDRGFDSALYSKSLGITDPKEIKETPPKYPIPLLVIVISVPIQMARAFALIGNSGESEIEPYGIRIIEPRAGIITNESASIIAKIKNEHOTQIVSOPVIAITIDMKKSTIQGTLANQGYTLMKKPKSDIAGSGTQNMADGVAIGYSPIYTAFWNGDQKSGISGIGTGLAGPGEKFMAYEHNKLLKRYVPAGIISIPVOAELGLPEELADKERIKNELISGTQGVKXSGMGEKMKLENNIYQIDIEINNDIENDTPEFEYELDNLSFNNNSINDNLSFNNNNNDSINDNEENKREDEIENLIEEPINLENKRPQDVLVNNNNDEMILIENTKEIRDEIVNETIETOSTELNSNNNEEKINNNDVGGEDIQDL"
5154. .5984
/gene="BB0733"
5154. .5984
/note="hypothetical protein; identified by Glimmer; putative"
codon_start=1
product="B. burgdorferi predicted coding region BB0733"
protein_id="AAC67086.1"
db_xref="pid:92688674"
/db_xref="gi:2688674"

gene
CDS
complement(6188. .7201)
/gene="BB0734"
complement(6188. .7201)
/gene="BB0734"
/note="similar to SP:P39153 PID:556881 GB:AL009126 percent identity: 37.80; identified by sequence similarity;
putative
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAC67081.1"
/db_xref="PID:g2688669"
/db_xref="GI:2688669"
/translation="MISTEIISSQIQAKKLIMGELVVFPEITYGAMANYNEDA
VAMILRGLPINNPLIYAVDTVKRIKELSEYIPKSAALMKRSPGLTYVLKSK
ISREFSGMLDVAIVRIIPANKATLILKASKPIYAPANSIKRSPSPFMALKEING
LVRGILKREKNKDNIGICESTGVGFDKDNVLLIPGATIKKMIENDELOKRYVNAE
TKMELEKPGNIEIKRKPIPYVLFKQDNIRRLANDQTLITKATLKYSLNPFNP
KNVIEFPEIAONLTKELVNSENNYKQILSEFLDEELGHSINNRKASSNPRI
NKK"
7268. .8095
/gene="BB0735"
7268. 8095
/gene="BB0735"
/note="similar to GB:L42023 PID:1004160 PID:1221934
PID:1204288 SP:Q57092 percent identity: 35.14; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="rare lipoprotein A (rLpA)"
/protein_id="AAC67080.1"
/db_xref="PID:g2688668"
/db_xref="GI:2688668"
/translation="MTEICFYIILINMRNLIALLDNKNFVLTFVFFIASHLSA
TVGLASWGEAFHGKTRNGEKEPDMAALTAHKELEPTNTYKVTNLNRSVYVRIND
RGPEFKDIDILSKIAEKLDLIGVAPKIEVTESSNEKRBSVSKDFKTFNIP
GVKEKKRIVNDSIVNKRKSLDLADYSADLEDTFYIOVGSYRKKDADRAYRL
KKAGLEFVVNSHGEPFYTFEPTNADYQKNIELISAGYKDTLIRKTVGESLMD"
8147. .8290
/gene="BB0736"
8147. 8290
/gene="BB0736"
/note="hypothetical protein; identified by Glimmer;
putative"
/codon_start=1
/transl_table=11
/product="B. burgdorferi predicted coding region BB0736"
/protein_id="AAC67085.1"
/db_xref="PID:g2688673"
/db_xref="GI:2688673"
/translation="MNSFINGPFFVFLYFNKFNVIINGNLNLDCDTLKFYFYCYV
VTL"
8269. .9276
/gene="BB0737"
8269. 9276
/gene="BB0737"
/note="similar to PID:699230 SP:Q50049 percent identity:
27.80; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="histidine phosphokinase/phosphatase, putative"
/protein_id="AAC67079.1"
/db_xref="PID:g2688667"
/db_xref="GI:2688667"
/translation="MLCSYFMKFLFDIPLPIMLIAPEMDVDTVFERNLHLIGSAGE
PHIYTEDLISKRGINGSKQSVQHVFLKPNLEINPLIAQIIGWGVPEGYAIDILGSM
GFWGIDIMNGPKSKIIKKGYCSALINNRSLAKETIILASKEACAREMLPLSYTRGCF

```

gene
9280. 11907
/gene="BB0738"
CDS
9280. 11907
/gene="BB0738"
/note="Similar to GB:M16318 SP:P11931 PID:143797 percent
identity: 47.36; Identified by sequence similarity;

Query Match
Best Local Similarity 94.4% Pred. No.3e+02; Length 11985;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTGTAAATTCCTAATGCTG 20
||||| ||||| |||
Db 10789 CTGTAAATTCCTAATGCTG 10772

RESULT 12
HS424L16 107484 bp DNA PRI 03-JUN-1999
LOCUS Human DNA sequence *** SEQUENCING IN PROGRESS *** from clone
424L16, complete sequence.
ACCESSION AL034373
NID 95002627
VERSION AL034373.L3 GI:5002627
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 107484)
TITLE Mashreghi-Mohammadi M.
JOURNAL Direct Submission
Submitted (03-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
hmqqueries@anger.ac.uk Clone requests: clonerequests@anger.ac.uk
On Jun 7, 1999 this sequence version replaced gi:4584767.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments.

FEATURES
source . Location/Qualifiers
1..107484
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="424L16"
BASE COUNT 33899 a 18771 c 20171 g 34643 t
ORIGIN

Query Match
Best Local Similarity 94.4% Pred. No.2.2e+02; Length 107484;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTGTAAATTCCTAATGCTG 20
||||| ||||| |||||
Db 11101 CTGTAAATTCCTAATGCTG 11118

RESULT 13
AC004098 166978 bp DNA PRI 01-MAY-1998
LOCUS Homo sapiens chromosome 17, clone HCT1305D20, complete sequence.
ACCESSION AC004098
NID 93097872
VERSION AC004098.1 GI:3097872;

```

	HFG.	human.
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE	1 (bases 1 to 166978)	
AUTHORS	Birren,B., Fasman,K., Linton,L., Nusbaum,C. and Lander,E.	
JOURNAL	Homo sapiens chromosome 17, clone HcIT305D20 unpublished	
REFERENCE	2 (bases 1 to 166978)	
AUTHORS	Birren,B., Fasman,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Batra,N., Beckery,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cooke,P., Daly,M.J., DePatre,E., Devon,K., Dewar,K., Durett,B., Edenfeldt,S., Ferrante,P., Forrest,C., Gage,D., Gardyna,S., Genshlmer,S., Geraldsgy,K., Gilmarlin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hul.L., Jacotot,L., Linton,L., MacKenzie,J., Margulis,N., McEwan,P., McGurt,A., Meldrim,J., Mollis,M., Morris,W., Morrow,J., Nachman,A., Naylor,I., O'Connor,T., Pavlin,B., Peterson,K., Ranganathan,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shyam,R., Sochoo,S., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Ye,W.J., Zemtseva,I., Zhao,J. and Zody.M.	
TITLE	Direct Submission	
JOURNAL	Submitted (01-FEB-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
REFERENCE	3 (bases 1 to 166978)	
AUTHORS	Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Batra,N., Beckery,R., Benn,J., Boatin,C., Boutwell,C., Brown,A., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P., Depyre,E., Devon,K., Dewar,K., Donelan,L., Durette,B., Edenfeldt,S., Ferreira,P., Fitzlugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Genshlmer,S., Geraldsgy,K., Gilmarlin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hul.L., Jacotot,L., Kann,L., MacDonald,P., Margulis,N., McEwan,P., McGurt,A., McKernan,K., Meldrim,J., Mollis,M., Morris,W., Morrow,J., Mychaleckyj,J., Nachman,A., Nahf,R., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody.M	
TITLE	Direct Submission	
JOURNAL	Submitted (01-MAY-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA	
COMMENT	On May 1, 1998 this sequence version replaced gi:3095021. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html .	
FEATURES	Location/Qualifiers	
SOURCE	1..166978	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="HCIT305D20"	
	/clone_1lb=Research Genetics/Cal Tech CITB-HSP-C (plates 195-384)"	
	/map="17"	
	/chromosome="17"	
	complement(727..1169)	
	/rpt_family="HTTC"	
	complement(1788..1893)	
	/rpt_family="LIMC4"	
	1934..1978	
	/rpt_family="CAN"	
	complement(2059..2357)	
	/rpt_family="AUSC"	
	complement(3360..3660)	
	/rpt_family="AUSX"	
	complement(3854..4144)	
	/rpt_family="AUSg"	
	complement(4191..4414)	
	/rpt_family="Aluud"	
	complement(4433..4621)	

```

repeat_region /rpt_family="L1ME"
complement(4741..5042)
repeat_region /rpt_family="Alusp"
5273..5569
repeat_region /rpt_family="Alusx"
6359..6391
repeat_region /rpt_family="AT_r1ch"
complement(6650..6707)
repeat_region /rpt_family="(CA)n"
complement(7730..8031)
repeat_region /rpt_family="Aluy"
8784..9067
repeat_region /rpt_family="Aluub"
complement(9222..9255)
repeat_region /rpt_family="AT_r1ch"
complement(9256..9554)
repeat_region /rpt_family="Alusx"
complement(9561..9685)
repeat_region /rpt_family="Alusq/x"
complement(9686..9764)
repeat_region /rpt_family="MIR"
complement(9764..9903)
repeat_region /rpt_family="L2"
10276..10360
repeat_region /rpt_family="L2"
10410..10582
repeat_region /rpt_family="MIR"
complement(11785..11874)
repeat_region /rpt_family="MER3"
complement(11881..1236)
repeat_region /rpt_family="L1MD3"
12578..12621
repeat_region /rpt_family="AT_r1ch"
complement(13069..13199)
repeat_region /rpt_family="L1MA7"
13238..13280
repeat_region /rpt_family="AT_r1ch"
14239..14330
repeat_region /rpt_family="AT_r1ch"
complement(15960..16041)
repeat_region /rpt_family="AT_r1ch"
complement(16384..16446)
repeat_region /rpt_family="AT_r1ch"
17101..17400
repeat_region /rpt_family="Aluy"
complement(19091..19129)
repeat_region /rpt_family="AT_r1ch"
19192..19288
repeat_region /rpt_family="AT_r1ch"
19651..19766
repeat_region /rpt_family="Aluub"
complement(19767..19814)
repeat_region /rpt_family="AT_r1ch"
complement(20491..20712)
repeat_region /rpt_family="MIR"
complement(20819..20890)
repeat_region /rpt_family="MER20"
21092..21127
repeat_region /rpt_family="AT_r1ch"
complement(21616..21748)
repeat_region /rpt_family="MER20"
22002..22031
repeat_region /rpt_family="AT_r1ch"
22541..22399
repeat_region /rpt_family="L2"
23656..23730
repeat_region /rpt_family="L2"
complement(24469..24780)
repeat_region /rpt_family="Aluub"
24847..24877
repeat_region /rpt_family="AT_r1ch"
complement(27965..28156)
repeat_region /rpt_family="L1ME"

```

```

repeat_region 28669..28720
repeat_region /rpt_family="L2"
repeat_region complement(28816..28870)
repeat_region /rpt_family="POLY_A"
repeat_region complement(28934..29213)
repeat_region /rpt_family="Alusx"
29355..29358
repeat_region /rpt_family="MIR"
repeat_region complement(29583..29628)
repeat_region /rpt_family="AT_r1ch"
31312..31370
repeat_region /rpt_family="MER47A"
complement(31362..31438)
repeat_region /rpt_family="MER47B"
complement(32276..32458)
repeat_region /rpt_family="Alusq/x"
complement(32461..32760)
repeat_region /rpt_family="Aluy"
complement(32808..33024)
repeat_region /rpt_family="Aluub"
33969..34139
repeat_region /rpt_family="MER5A"
34934..35227
repeat_region /rpt_family="Aluub"
35433..35718
repeat_region /rpt_family="Alusx"
35719..35768
repeat_region /rpt_family="AT_r1ch"
complement(37095..37519)
repeat_region /rpt_family="L1MA4A"
complement(37903..38084)
repeat_region /rpt_family="MIR"
38440..39130
repeat_region /rpt_family="L2"
39405..39567
repeat_region /rpt_family="Aluub"
42638..42663
repeat_region /rpt_family="AT_r1ch"
complement(44243..44397)
repeat_region /rpt_family="L1PB2"
complement(44447..44658)
repeat_region /rpt_family="L1PB2"
44669..44749
repeat_region /rpt_family="(TA)n"
complement(44757..44838)
repeat_region /rpt_family="Aluub"
45894..46277
repeat_region /rpt_family="MLT2CB"
complement(46281..46325)

```

Query Match 82.0%; Score 16.4; DB 11; Length 166978;
 Best Local Similarity 94.4%; Pred. No. 2.1e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 CTGTAATTCATGCG 20
 Db 164154 CTGTAATTCATGCG 164171

RESULT 14
 AC005968/c AC005968 145603 bp DNA PRI 22-NOV-1998
 LOCUS Homo sapiens chromosome 18, clone hRPK.24_A.23, complete sequence.
 DEFINITION AC005968
 ACCESSION AC005968
 NID 93907448
 VERSION AC005968.1 GI:3907448
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 145603)
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.

```

TITLE      Homo sapiens chromosome 18, clone hRPK.24_A.23
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 145603)
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,K.,
            Gage,D., Gardyna,S., Geradigery,K., Grant,G., Hagos,B., Heathford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
            McQuirk,A., McKernan,K., Meldrum,J., Molla,M., Morris,M., Morrow,J.,
            Mychaleckyj,J., Nafé,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Testfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            *Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.

TITLE      Direct Submission
JOURNAL    Submitted (14-NOV-1998) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baker,J., Baldwin,J., Barna,N., Beckert,R., Benn,J., Boutwell,C.,
            Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
            Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,K.,
            Gage,D., Gardyna,S., Geradigery,K., Grant,G., Hagos,B., Heathford,A.,
            Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
            Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
            McQuirk,A., McKernan,K., Meldrum,J., Molla,M., Morris,M., Morrow,J.,
            Mychaleckyj,J., Nafé,R., Naylor,J., Niloff,M., O'Connor,T.,
            O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
            Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
            Subramanian,A., Testfaye,S., Tichovolsky,N., Torruella-Miller,I.,
            Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
            Ye,W.J., Zhao,J. and Zody,M.

TITLE      Direct Submission
JOURNAL    Submitted (22-NOV-1998) Whitehead Institute/MIT Center for Genome
REFERENCE  Research, 320 Charles Street, Cambridge, MA 02141, USA
AUTHORS    On Nov 22, 1998 this sequence version replaced gi:3873178.
            All repeats were identified using RepeatMasker. Smit, A.F.A. &
            Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES   Location/Qualifiers
            source          1..145603
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="hRPK.24_A.23"
                        /clone_lib="RPCT-II human BAC library"
                        /map="18"
                        /chromosome="18"
            repeat_region   336..370
                        /rpt_family="AT_rich"
                        complement(485..526)
            repeat_region   /rpt_family="polypurine"
                        758..838
            repeat_region   /rpt_family="L2"
                        complement(862..1239)
            repeat_region   /rpt_family="L1"
                        1673..1970
            repeat_region   /rpt_family="AluYb8"
                        2764..3334
            repeat_region   /rpt_family="L2"
                        3354..3717
            repeat_region   /rpt_family="THE1C"
                        3720..5186
            repeat_region   /rpt_family="THE1C-internal"
                        5302..5482
            repeat_region   /rpt_family="THE1C"
                        complement(5484..5844)
            repeat_region   /rpt_family="MSTD"
                        complement(6351..6650)
            repeat_region   /rpt_family="AluY"
                        6770..7128
            repeat_region   /rpt_family="L2"
                        complement(747..7870)
            repeat_region   /rpt_family="L2"
                        complement(8028..8063)
            repeat_region   /rpt_family="(CA)n"
                        8457..8641
            repeat_region   /rpt_family="MER5A"
                        complement(9019..9242)
            repeat_region   /rpt_family="L1ME3A"
                        9627..9747
            repeat_region   /rpt_family="purine-rich"
                        9865..10123
            repeat_region   /rpt_family="L2"
                        10160..10274
            repeat_region   /rpt_family="MIR"
                        10925..10959
            repeat_region   /rpt_family="AT_rich"
                        complement(11325..11654)
            repeat_region   /rpt_family="AluSp"
                        complement(12565..12889)
            repeat_region   /rpt_family="L2"
                        13885..14185
            repeat_region   /rpt_family="AluY"
                        complement(14214..14286)
            repeat_region   /rpt_family="polypurine"
                        14790..14817
            repeat_region   /rpt_family="AT_rich"
                        complement(14984..15281)
            repeat_region   /rpt_family="AluSg"
                        complement(16483..16863)
            repeat_region   /rpt_family="THE1C"
                        16922..16960
            repeat_region   /rpt_family="AT_rich"
                        complement(18094..18322)
            repeat_region   /rpt_family="MER50"
                        complement(18323..18630)
            repeat_region   /rpt_family="AluY"
                        complement(18631..19059)
            repeat_region   /rpt_family="MER50"
                        complement(22147..22442)
            repeat_region   /rpt_family="AluSg"
                        complement(22767..23083)
            repeat_region   /rpt_family="AluSx"
                        complement(23395..23896)
            repeat_region   /rpt_family="MER67D"
                        complement(24797..24817)
            repeat_region   /rpt_family="(TAA)n"
                        26138..26459
            repeat_region   /rpt_family="MER57A"
                        26791..26823
            repeat_region   /rpt_family="AT_rich"
                        27092..27118
            repeat_region   /rpt_family="AT_rich"
                        27272..27415
            repeat_region   /rpt_family="L1ME"
                        27476..27538
            repeat_region   /rpt_family="L2"
                        28152..28447
            repeat_region   /rpt_family="AluSx"
                        28457..28680
            repeat_region   /rpt_family="L2"
                        29046..29113
            repeat_region   /rpt_family="L1MC4"
                        complement(29352..29455)
            repeat_region   /rpt_family="Char1el"
                        complement(29469..29678)
            repeat_region   /rpt_family="Char1el"
                        complement(29931..29963)
            repeat_region   /rpt_family="(CA)n"
                        32899..32924
            repeat_region   /rpt_family="(TA)n"

```

```

repeat_region complement(33041..33537)
repeat_region /rpt_family="MUT2B"
repeat_region 33731..33855
repeat_region /rpt_family="MER5B"
repeat_region 33975..34069
repeat_region /rpt_family="LIME"
repeat_region complement(34102..34137)
repeat_region /rpt_family="(CA)n"
repeat_region 34138..34469
repeat_region /rpt_family="LIME"
repeat_region 34780..34810
repeat_region /rpt_family="(CAA)n"
repeat_region 34837..34871
repeat_region /rpt_family="POLY_A"
repeat_region 34889..35221
repeat_region /rpt_family="LIME"
repeat_region 35290..35679
repeat_region /rpt_family="LIME2"
repeat_region 35680..35956
repeat_region /rpt_family="AluSg1"
repeat_region 35957..35979
repeat_region /rpt_family="(CAA)n"
repeat_region 35981..36581
repeat_region /rpt_family="LIME2"
repeat_region 36699..37964
repeat_region /rpt_family="LIME2"
repeat_region complement(37965..38274)
repeat_region /rpt_family="AluY"
repeat_region 38275..38836
repeat_region /rpt_family="LIME2"
repeat_region complement(38858..39349)
repeat_region /rpt_family="L2"
repeat_region 39533..39691
repeat_region /rpt_family="MIR"
repeat_region complement(40792..41086)
repeat_region /rpt_family="MER33"
repeat_region complement(41227..41324)
repeat_region /rpt_family="L2"
repeat_region 42393..42450
repeat_region /rpt_family="AT-rich"
repeat_region complement(42660..42772)
repeat_region /rpt_family="LTR40s"

Query Match      82.0%; Score 16.4; DB 11; Length 145603;
Best Local Similarity 94.4%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 CTGTAATCTAATGCTG 20
        |||||||
Db 103680 CTTTAATCTAATGCTG 103663

RESULT 15
G09637/c      387 bp      DNA      STS      14-ADG-1995
LOCATION      human STS CHLC.GATA47F01.P15242 clone GATA47F01.
ACCESSION      G09637
NID      9941486
G09637.1 GI:941486
KEYWORDS      STS sequence; primer; sequence tagged site.
SOURCE      human vector-pUC1 host-E.coli dutung+ (DH10B) Marker selected
              genomic DNA prepared from X1 individual of French nationality.
ORGANISM      Homo sapiens
              Eukaryota; Eukaryotes; Metazoa; Chordata;
              Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata;
              Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates;
              Catarrhini; Homidae; Homo.
              1 (bases 1 to 387)
              Murray,J., Sheffield,V, Weber,J.L., Duyk,G. and Buetow,K.H.
              Cooperative Human Linkage Center
              Unpublished (1995)
              Synonyms: GATA47F01, CHLC.GATA47F01.T15166
              COMMENT
              GDB: G00-365-330

```

```

Contact: Dr. Jeffrey C. Murray
Dotti
The University of Iowa
Department of Pediatrics, Iowa City, IA 52242, USA
Tel: (319) 356-3508
Fax: (319) 356-3347
Email: jelf-murray@uiowa.edu

Primer A: TATGAGCTAATGCCTTACAGTAA
Primer B: CAGGAAATAGACACACAG
STS size: 117

PCR Profile:
denature: 30 seconds at 94 degrees C
annealing: 75 seconds at 55 degrees C
extension: 15 seconds at 72 degrees C
PCR cycles: 27
extension: 6 minutes at 72 degrees C

Protocol:
Template: 30ng genomic DNA
Primer: each 1.5 pmole
dNTPs: each 200 uM
Tag Polymerase: 0.3 units
Total Vol: 10 uL

Buffer:
MgCl2: 1.5mM
KCl: 50mM
Tris: 10mM
pH: 8.3

Equivalent set: GATA47G01

FEATURES
source      Location/Qualifiers
            1..387
            /organism="Homo sapiens"
            STS      233..349
            primer_bind 233..256
            primer_bind complement(329..349)
BASE COUNT 99 a 76 c 74 g 126 t 12 others
ORIGIN

Query Match      82.0%; Score 16.4; DB 14; Length 387;
Best Local Similarity 94.4%; Pred. No. 4.8e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCTGTAATCTAATGC 18
        |||||||
Db 119 TGCTGTAATCTAATCC 102

```

Search completed: September 13, 1999, 15:55:35
Job time: 4563 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:31 ; Search time 255.05 Seconds

(without alignments)
19.619 Million cell updates/sec

Title: US-09-325-095-21

Perfect score: 20

Sequence: 1 TCGCTGAATTCCTAATGCTG 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	20	100.0	20	059020	Antisense PCR prim
2	20	100.0	3498	057012	Ptdins 3-kinase 11
3	16.4	82.0	110000	X20248_01	Continuation (2 of
4	15.8	79.0	1797	004597	Equine hemagglutinin
5	15.8	79.0	1788	N71067	Sequence encoding
6	15.8	79.0	1762	029111	EYV HA (A1/Pontain
7	15.8	79.0	1762	029112	EYV HA (A2/Suffolk
8	15.8	79.0	1698	V49391	EYV Fontainebleau
9	15.8	79.0	8395	X13154	Enterococcus faeca
10	15.4	77.0	53585	X20251	Borrelia burgdorfe
11	15.2	76.0	226	020447	B.burgdorferi str
12	15.2	76.0	226	020448	B.burgdorferi str
13	15.2	76.0	2976	T04687	Black widow spider
14	15.2	76.0	3706	T04688	Black widow spider
15	15.2	76.0	4278	V52286	Streptococcus pneu
16	15.2	76.0	110000	V21209_15	Continuation (16 o
17	15.2	76.0	5910	V66798	Rice bacterial lea
18	15.2	76.0	13856	V74342	Staphylococcus aur
19	15.2	76.0	17764	X13238	Enterococcus faeca
20	15.2	76.0	31517	X13117	Enterococcus faeca
21	15.2	76.0	1014	X20074	Enterococcus faeca
22	15.2	76.0	892	X20075	Enterococcus faeca
23	15.2	76.0	110000	X20248_01	Continuation (2 of
24	15.2	76.0	21170	X20535	Polynucleotide seq
25	14.8	74.0	1189	071003	Clone V51 hybridiz
26	14.8	74.0	1512	V39970	Saccharomyces cere
27	14.8	74.0	3930	V39971	Saccharomyces cere
28	14.8	74.0	21338	V52153	Streptococcus pneu
29	14.8	74.0	1267	V65201	DNA encoding a S.
30	14.8	74.0	6146	X13076	Enterococcus faeca
31	14.8	74.0	2639	X52226	Protein PRO220 CDN
32	14.4	72.0	350	T26581	Human gene signatu
33	14.4	72.0	2558	T59360	Mouse Fas-associat
34	14.4	72.0	3833	V17116	Metabotropic gluta
35	14.4	72.0	3321	V04206	Human metabotropic
36	14.4	72.0	3321	V04207	Human metabotropic
37	14.4	72.0	3567	X13164	Enterococcus faeca
38	14.4	72.0	251	X12586	Human diallelic po
39	14.4	72.0	110000	X20248_03	Continuation (4 of
40	14.2	71.0	761	N80472	Sequence specifica
41	14.2	71.0	1052	N80475	Genomic form (7PB1
42	14.2	71.0	1052	N82441	Genomic form (7PB7
43	14.2	71.0	1052	N82442	Genomic form (7PB7

ALIGNMENTS

c 44 14.2 71.0 8937 1 020602
45 14.2 71.0 1845 1 022614

Nfi gene, DNA sequ
R. Communis syntha

RESULT 1
ID 059020 standard; DNA; 20 BP.

AC 059020:

DE 12-APR-1994 (first entry)

DE Antisense PCR primer for p110.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ss; amplification; p110.

OS Synthetic.

PN W09321328-A.

PD 28-OCT-1993.

PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Dhond R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parker PJ, Volinia S, Waterfield MD;

PI WPI; 93-351738/44.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity; useful for controlling cell proliferation

PS Example 1; Page 41; 146pp; English.

CC An SGARF-1 cell line was established by transfection of bovine

CC adrenal cortex zona fasciculata cells with pSVneo. For p110

CC poly (A)+ RNA was reversed transcribed with rth DNA polymerase and the

CC antisense primer shown. For DNA polymerisation a sense primer was added

CC and PCR performed. The prod. was p110.

CC See also 051155-6, 059012-23 and 057522-3.

SQ Sequence 20 BP: 5 A; 3 C; 4 G; 8 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.48;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCGCTGAATTCCTAATGCTG 20

DB 1 TCGCTGAATTCCTAATGCTG 20

RESULT 2

ID 057012/c

AC 057012: standard; cDNA to mRNA; 3498 BP.

DE 31-AUG-1994 (first entry)

DE Ptdins 3-kinase; 110 kD catalytic subunit cDNA.

KW 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;

KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;

KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;

OS Bos taurus.

FN Key

FT Cds

FT 1..3207

FT /*tag- a

FT /product- p110

PN W09403609-A.

PD 17-FEB-1994.

PF 05-AUG-1993; G01651.

PR 05-AUG-1992; GB-016654.

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;

DR WPI; 94-065697/08.

DR P-PSDB; R46294.

PT Eukaryotic cells transformed with mammalian phospholipid or

PT protein kinase DNA; useful in assays for compounds involved in

PT cell growth regulation and for treating cancers

PS Disclosure: Fig 1: 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylinositol (PtdIns) 3-kinase. This sequence was transformed into Schizosaccharomyces pombe cells under the regulatory control of the CC promoter in an embodiment of the invention. In the presence of CC thiamine the promoter is inactive and the cells carrying the PtdIns CC catalytic subunit plasmid grow as the parental strain. In the absence of thiamine the mnt promoter functions and the PtdIns 3-kinase CC catalytic subunit is induced. PtdIns activity is substantially increased under these conditions. Cells containing constructs such as CC this, are useful in assays for detecting compounds involved in cell growth regulation. It is also used as the basis for detecting CC compounds for treating cancers and the formation of blood vessel CC plaques.
 SQ Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 100.0%; Score 20; DB 1; Length 3498;
 Best Local Similarity 100.0%; Pred. No. 0.66; Mismatches 0; Indels 0; Gaps 0;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTG 20
 |||||
 DB 3325 TGCTGTAATTCATGCTG 3306

Query Match 82.0%; Score 16.4; DB 1; Length 110000;
 Best Local Similarity 94.4%; Pred. No. 39; Mismatches 1; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 CTGTAATTCATGCTG 20
 |||||
 DB 30454 CTGTAATTCATGCTG 30471

RESULT 4
 ID 004597 standard; DNA; 1797 BP.
 AC 004597;
 DT 10-MAR-1993 (revised)
 DT 02-OCT-1990 (first entry)
 DE Equine haemagglutinin H3 (EIV-A2).
 KW Recombinant vaccines; equine influenza virus; haemagglutinin; H3; KW neuraminidase; N8; ss.
 OS Equine influenza virus.
 FH Key
 FT cds
 FT Location/Qualifiers
 FT 54..1748
 FT /tag- a
 FT /Product-EIV-A2 haemagglutinin H3
 FT 102..1182
 FT /tag- b
 FT /Product-N-terminal HA1 50KD portion
 FT 1186..1748
 FT /tag- c
 FT /Product-C-terminal HA2 27KD portion
 PN US4920213-A.

PD 24-APR-1990
 PF 21-JUL-1986; 888250.
 PR 20-JUN-1985; US-747020.
 PR 21-JUL-1986; US-888250.
 PA (BIOT-) Biotech Res Partner.
 PI Dale B. Cordell B.
 DR WPI: 90-163647/21.
 DR P-PSDB; R04943.
 PT Recombinant vaccines against equine influenza virus - produced using DNA PT sequences encoding haemagglutinin and neuraminidase glycoprotein(s).
 PS Disclosure: 27pp; English.
 CC There are nine neuraminidase (NA) subtypes and twelve haemagglutinin (HA) CC subtypes. The strain carrying H3N8 glycoproteins is designated equine CC influenza virus (EIV)-A2. The cDNA sequences encoding these CC will be useful in the construction of diagnostic probes for the CC disease and of probes for obtaining new cDNAs of the mutated form of CC the virus. Recombinant vaccines are produced.
 SQ See also 004596-004599.
 SQ Sequence 1797 BP; 639 A; 341 C; 413 G; 404 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1797;
 Best Local Similarity 89.5%; Pred. No. 57; Mismatches 2; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19
 |||||
 DB 464 TGCTGTAATTCATGCT 446

Query Match 79.0%; Score 15.8; DB 1; Length 1797;
 Best Local Similarity 89.5%; Pred. No. 57; Mismatches 2; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19
 |||||
 DB 464 TGCTGTAATTCATGCT 446

RESULT 5
 ID N71067 standard; DNA; 1788 BP.
 AC N71067;
 DT 08-MAR-1991 (first entry)
 DE Sequence encoding equine influenza virus strain H3N8 (EIV-A2)
 DE haemagglutinin protein H3.
 KW HA; vaccine; vaccina; ds.
 OS Equine influenza virus.
 FH Key
 FT cds
 FT Location/Qualifiers
 FT 54..1748
 FT /tag- a
 FT /tag- b
 FT /label- HA 1
 FT 1086..1748
 FT /tag- c
 FT /label- HA 2
 PN W08607593-A.
 PD 31-DEC-1986.
 PE 20-JUN-1986; U01343.
 PR 20-JUN-1985; US-747020.
 PA (BIOT-) BIOTECHN RES PARTNER.
 PI Dale B. Cordell B.
 DR WPI: 87-007191/01.
 DR P-PSDB; P70711.
 PT Preventing equine influenza virus infection - using recombinant PT vaccines produced using DNA sequences encoding haemagglutinin and PT neuraminidase glycoproteins
 PS Disclosure: Fig 2: 63pp; English.
 CC Peptides derived from the haemagglutinin H7 and H3 and neuraminidase CC N7 and N8 genes may be used to derive antigenic peptides useful in CC vaccination against equine influenza virus infection.
 CC Abs raised to the peptides may be used in diagnosis of the infection CC and construction of probes to mutated forms of the virus.
 SQ Sequence 1788 BP; 637 A; 340 C; 409 G; 402 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1788;
 Best Local Similarity 89.5%; Pred. No. 57; Mismatches 2; Indels 0; Gaps 0;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19

Db 464 TGCCTGATCTATGTT 446

RESULT 6

ID 029111/c

029111 standard; DNA: 1762 BP.

AC 029111; 24-FEB-1993 (first entry)

DE EIV HA (A2/Fontainebleau/79).

KW Equine influenza virus; EIV; haemagglutinin; HA: A1/Fontainebleau/79;

KW expression cassette; NYVAC; ALVAC; recombinant vector;

KW polymerase chain reaction; PCR; vaccinia virus; H6 promoter;

KW canarypox virus; Copenhagen vaccine strain; virulence factor;

KW deletion loci; recipient loci; ss.

OS Synthetic.

PN WO9215672-A.

PD 17-SEP-1992.

PF 09-MAR-1992; 001906.

PR 07-MAR-1991; US-666056.

PR 11-JUN-1991; US-713967.

PA 06-MAR-1992; US-847951.

PI (VIR-) VIROGENETICS CORP.

PI Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP,

PI Limbach KJ, Norton EK, Paolletti E, Perkins ME, Pincus SE,

PI Riviere M, Tartaglia J, Taylor J;

PI WPI: 92-331718/40.

PT Vaccine comprises recombinant, attenuated pox-virus - use for

PT vaccinating against viral infections such as rabies, hepatitis B,

PS HIV, HSV, EBV, CMV, mumps etc.

PS Disclosure: Fig 24: 456pp; English.

CC The sequence given encodes the equine influenza virus (EIV)

CC haemagglutinin (HA) (A1/Fontainebleau/79). This sequence was used to

CC generate an expression cassette for the insertion of the EIV HA

CC gene into NYVAC and ALVAC recombinant vectors. The HA gene sequence

CC was isolated from an EIV cDNA library and was amplified by polymerase

CC chain reaction. The HA gene sequence was fragmented and then

CC reconstituted aligned with the vaccinia virus H6 promoter. NYVAC is

CC derived from a Copenhagen vaccine strain of vaccinia virus and ALVAC

CC is derived from a canarypox virus which has been modified by deletion

CC of non-essential regions of the genome encoding known or potential

CC virulence factors. The deletion loci of both vectors were engineered

CC as recipient loci for the insertion of foreign genes. See also

CC Q35501-864.

PR 06-MAR-1992; US-847951.

PA (VIR-) VIROGENETICS CORP.

PI Cox WI, De Taisne C, Francis J, Gettig RR, Johnson GP,

PI Limbach KJ, Norton EK, Paolletti E, Perkins ME, Pincus SE,

PI Riviere M, Tartaglia J, Taylor J;

PI WPI: 92-331718/40.

PT Vaccine comprises recombinant, attenuated pox-virus - use for

PT vaccinating against viral infections such as rabies, hepatitis B,

PS HIV, HSV, EBV, CMV, mumps etc.

PS Disclosure: Fig 25: 456pp; English.

CC The sequence given encodes the equine influenza virus (EIV)

CC haemagglutinin (HA) (A2/Suffolk/89). This sequence was used to

CC generate an expression cassette for the insertion of the EIV HA

CC gene into NYVAC and ALVAC recombinant vectors. The HA gene sequence

CC was isolated by polymerase chain reaction from an M13 clone. Some

CC non-conserved base-changes were observed due to the amplification

CC process. The EIV HA gene was linked to the 13L promoter and inserted

CC into a vaccinia insertion plasmid deleted for ORF's C6L-K1L. NYVAC is

CC derived from a Copenhagen vaccine strain of vaccinia virus and ALVAC

CC is derived from a canarypox virus which has been modified by deletion

CC of non-essential regions of the genome encoding known or potential

CC virulence factors. The deletion loci of both vectors were engineered

CC as recipient loci for the insertion of foreign genes. See also

CC Q35501-864.

CC Sequence 1762 BP; 637 A; 328 C; 384 G; 412 T;

SQ

Query Match 79.0%; Score 15.8; DB 1; Length 1762;

Best Local Similarity 89.5%; Pred. No. 57;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTATGCT 19

Db 440 TGCCTGAATCTATGCT 422

RESULT 8

ID 029111/c

029111 standard; DNA: 1698 BP.

AC 029111; 28-OCT-1998 (first entry)

DE EIV Fontainebleau strain haemagglutinin gene.

KW Multivalent vaccine; Horse; pathogen; respiratory disease; EHV; EIV;

KW Clostridium tetani; Borrelia burgdorferi; equine influenza virus; EEEV;

KW Eastern equine encephalomyelitis virus; equine herpesvirus; WEEV; VEEV;

KW Western equine encephalomyelitis virus; digestive disease; rabies virus;

KW Venezuelan equine encephalomyelitis virus; vector; primer; PCR;

KW amplification; haemagglutinin; ss.

OS Equine influenza virus.

PN Key

PD 1.1698

PF 1.1698

PA (INMR) RHONE MERIEUX SA.

PI Audonnet JCF, Bouchardon A, Riviere MEA.

PI WPI: 98-112826/11.

PT P-PSDB: W44946.

PT Multi-valent polynucleotide vaccines against equine pathogens

PT consist of at least 3 plasmids able to express protective antigens

PT from specified viruses

PS Example 14; Fig 8; 49pp; French.

CC The invention relates to a multivalent vaccine for protecting horses

CC against several pathogens, especially pathogens associated with

CC respiratory and digestive diseases. The pathogens are especially

CC selected from equine herpesvirus (EHV), equine influenza virus (EIV),

CC Clostridium tetani, Borrelia burgdorferi, Eastern, Western or Venezuelan

CC equine encephalomyelitis viruses (EEEV, WEEV and VEEV, respectively) and

CC rabies virus. The vaccines are preferably composed of polynucleotide

CC sequences encoding 3 antigens, all as part of vectors.

CC This sequence represents the coding region of the EIV Fontainebleau
CC strain haemagglutinin gene. The sequence was subcloned into the plasmid
CC pVR1012 to generate the plasmid PAB099 for use in the vaccine.
CC
CC Sequence 1698 BP; 604 A; 319 C; 384 G; 391 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1698;
Best Local Similarity 89.5%; Pred. No. 57;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19
||||| |||||||
DB 411 TGCTGTAATTCATGCT 393

RESULT 9
X13154 standard; DNA: 8395 BP.
AC X13154;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:217.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN MO9850555-A2.
PD 12-NOV-1998.
PE 04-MAY-1998; 008985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI: 99-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1: Page 1103-1107; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X13154 to X13199 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism. In vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
CC Sequence 8395 BP; 2616 A; 1692 C; 1478 G; 2607 T;

Query Match 79.0%; Score 15.8; DB 1; Length 8395;
Best Local Similarity 89.5%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 19
||||| |||||||
DB 6161 TGCAAGTAATTCATGCT 6179

RESULT 10
X20251/c
ID X20251 standard; DNA: 53585 BP.
AC X20251;
DT 04-MAY-1999 (first entry)

DE Borrelia burgdorferi polynucleotide sequence #4.
KW Borrelia burgdorferi; spirochete; bacterium; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.

OS Borrelia burgdorferi.
PN WO9858943-A1.
PD 30-DEC-1998.
PE 18-JUN-1998; 012764.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053844.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PI (MEDT-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.

PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1: Page 801-831; 1128pp; English.

CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
CC Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;

Query Match 77.0%; Score 15.4; DB 1; Length 53585;
Best Local Similarity 94.1%; Pred. No. 11e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCT 17
||||| |||||||
DB 11703 TGCTGTAATTCATGCT 11687

RESULT 11
O20447
ID O20447 standard; DNA: 226 BP.
AC O20447;
DT 15-APR-1992 (first entry)

DE B.burgdorferi strain P/B1 flagellin gene fragment.
KW tick-borne spirochete; Lyme disease; polymerase chain reaction; PCR;
KW Interprimer region; G1; G2; VS3; ss.
OS Borrelia burgdorferi.
PN WO9119814-A.
PD 26-DEC-1991.
PE 12-JUN-1991; 004190.
PR 15-JUN-1990; US-538957.
PR 25-APR-1991; US-691188.
PA (BAXT) BAXTER DIAGNOSTICS.
PI Picken RN, Ammons HC;
DR WPI: 92-024430/03.

PT New DNA primers useful in diagnosis of Lyme disease - comprise
PT complementary strand specifically flanking Borrelia burgdorferi
PT sequence
PS Disclosure: Fig 6; 47pp; English.

CC This sequence corresponds to nucleotides 517 to 742 of the
CC B.burgdorferi flagellin gene sequence of Q20720. The region was
CC amplified using primers flanking a region of relative non-homology
CC between B.burgdorferi and B.hemissi flagellin gene sequences. A
CC comparison of amplified sequences from a number of B.burgdorferi
CC strains (see Q20446-9) indicated several positions of mismatched
CC bases. By selecting a region containing minimally 5 mismatched
CC bases (i.e. between nucleotides 78 to 126), three probes were
CC obtained (see Q20720, Q21925-6) which could distinguish three
CC subgroups of B.burgdorferi in an assay in which the interprimer
CC region was first amplified and then probed. Other strains which
CC belong to the P/BI subgroup are G1, G2, VS3 (all with sequences
CC identical to P/BI) and VS185. See also Q20721, Q20723-5 and Q21927.
CC Sequence 226 BP; 65 A; 46 C; 53 G; 62 T;

Query Match 76.0%; Score 15.2; DB 1; Length 226;
Best Local Similarity 85.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTG 20
|||||
DB 39 TGCTGTAATTCATGCTG 58

RESULT 12
020448
ID Q20448 standard; DNA: 226 BP.
AC Q20448;
DT 15-APR-1992 (first entry)
DE B.burgdorferi strain VS185 flagellin gene fragment.
KW tick-borne spirochaete; Lyme disease; polymerase chain reaction; PCR;
KM Interprimer region; P/B1 subgroup; G1; G2; VS3; SS.
OS Borrelia burgdorferi.
PN W09119814-A.
PD 26-DEC-1991.
PE 12-JUN-1991; U04190.
PR 15-JUN-1990; US-538957.
PR 25-APR-1991; US-691188.
PA (BAXT) BAXTER DIAGNOSTICS.
PI Picken RN, Ammons HC.
PT WPI: 92-024430/03.
PT New DNA primers useful in diagnosis of Lyme disease - comprise complementary strand specificity flanking Borrelia burgdorferi PT sequence
PT Disclosure: Fig 6: 47pp; English.
TS This sequence corresponds to nucleotides 517 to 742 of the CC B.burgdorferi flagellin gene sequence of Q20720. The region was CC amplified using primers flanking a region of relative non-homology CC between B.burgdorferi and B.hemissi flagellin gene sequences. A CC comparison of amplified sequences from a number of B.burgdorferi CC strains (see Q20446-9) indicated several positions of mismatched CC bases (i.e. between nucleotides 78 to 126), three probes were CC obtained (see Q20720, Q21925-6) which could distinguish three CC subgroups of B.burgdorferi in an assay in which the interprimer CC region was first amplified and then probed. Strain VS185 belongs to CC the P/B1 subgroup, along with strains G11, G2, VS3.
CC See also Q20721, Q20723-5 and Q21927.
CC Sequence 226 BP; 67 A; 48 C; 52 G; 59 T;

Query Match 76.0%; Score 15.2; DB 1; Length 226;
Best Local Similarity 85.0%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTG 20
|||||
DB 39 TGCTGTAATTCATGCTG 58

RESULT 13
T04687
ID T04687 standard; DNA: 2976 BP.
AC T04687;
DT 16-MAR-1996 (first entry)
DE Black widow spider delta-latroinsectotoxin truncated gene DNA.
KW Black widow spider; insecticide; pesticide; delta-latroinsectotoxin;
KM Spider venom; neurotoxin; toxin; ds.
OS Latrodectus mactans tedeclimguttatus.
FH Key Location/Qualifiers
FT cds 1..2976
/*tag=a

PN GB2288807-A.
PD 01-NOV-1995.
PE 24-APR-1995; 008298.
PR 27-APR-1994; GB-008466.
PR 27-APR-1994; GB-008466.

PA (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
PI Bell DR, Dulubova I, Galkina TG, Grishin E, Khovtchey MV;
PI Krasnoperov V, Pluzhnikov KA, Shamotlenko OG, Usherwood PNR;
PI Volkova T, Galkina T, Khovtchey MV;
DR WPI: 95-360758/47.
DR P-PSDB: R80096.
PT Polypeptide(s) expressed by truncated genes, esp. spider
PT delta-latroinsectotoxin - also related non toxic precursor
PT polypeptide. Isolated from Black Widow spider, useful as insecticide
PS Claim 10; Page 31-35; 62pp; English.
CC This truncated gene may be expressed recombinantly in E. coli BL21 (CC DE3) cells transformed with pT7-7 vectors comprising the truncated CC form of the sequence. The encoded protein is an insect-specific CC neurotoxin, delta-latroinsectotoxin, which is useful as an oral or CC topical insecticide. Expressing the truncated gene allows large- CC scale production of active toxin in bacteria, eliminating the need CC to extract it from spider venom. This neurotoxin is harmless to CC mammals and plants.
SQ Sequence 2976 BP; 1089 A; 455 C; 571 G; 861 T;

Query Match 76.0%; Score 15.2; DB 1; Length 2976;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTG 20
|||||
DB 768 TGATGTAATTCATGCTG 787

RESULT 14
T04688
ID T04688 standard; DNA: 3706 BP.
AC T04688;
DT 16-MAR-1996 (first entry)
DE Black widow spider delta-latroinsectotoxin gene precursor DNA.
KW Black widow spider; insecticide; pesticide; delta-latroinsectotoxin;
KM Spider venom; neurotoxin; toxin; ds.
OS Latrodectus mactans tedeclimguttatus.
FH Key Location/Qualifiers
FT cds 45..3686
/*tag=a

PN GB2288807-A.
PD 01-NOV-1995.
PE 24-APR-1995; 008298.
PR 27-APR-1994; GB-008466.
PA (BRTE-) BRITISH TECHNOLOGY GROUP LTD.
PI Bell DR, Dulubova I, Galkina TG, Grishin E, Khovtchey MV;
PI Krasnoperov V, Pluzhnikov KA, Shamotlenko OG, Usherwood PNR;
PI Volkova T, Galkina T, Khovtchey MV;
DR WPI: 95-360758/47.
DR P-PSDB: R80097.
PT Polypeptide(s) expressed by truncated genes, esp. spider
PT delta-latroinsectotoxin - also related non toxic precursor
PT polypeptide. Isolated from Black Widow spider, useful as insecticide
PS Claim 62; Page 38-43; 62pp; English.
CC This gene encodes an insect-specific neurotoxin, delta- CC lactroinsectotoxin, which is useful as an oral or topical CC insecticide. The gene has been cloned into E. coli (HMS 174/ CC PT7.deltarl). This neurotoxin is harmless to mammals and plants.
CC Sequence 3706 BP; 1342 A; 588 C; 692 G; 1084 T;

Query Match 76.0%; Score 15.2; DB 1; Length 3706;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTG 20
|||||
DB 896 TGATGTAATTCATGCTG 915

RESULT 15

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

nm nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds

(without alignments)
23.280 Million cell updates/sec

Title: US-09-325-095-21

Perfect score: 20
Sequence: 1 TGCCTGAATCTAATGCTG 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	20	100.0	475	25	N44819	N44819 yY39a10.r1
C 2	20	100.0	343	31	AA298585	AA298585 EST114216
C 3	20	100.0	633	49	A1632172	A1632172 t865907.x
C 4	18.4	92.0	136	25	N84360	N84360 KR7998F Hum
C 5	18.4	92.0	340	34	AA508979	AA508979 MBAPCWH0
C 6	17.4	87.0	246	21	R10988	R10988 yF8C09.r1
C 7	16.8	84.0	580	36	C80053	C80053 C80053 Mous
C 8	16.8	84.0	208	43	A1102178	A1102178 EST211467
C 9	16	80.0	613	39	C83915	C83915 C83915 Dict
C 10	16	80.0	580	44	AU037810	AU037810 AU037810
C 11	15.8	79.0	496	22	R52778	R52778 Y996C11.r1
C 12	15.8	79.0	500	22	R59012	R59012 Y996C01.r1
C 13	15.8	79.0	214	25	N50767	N50767 YY90F02.r1
C 14	15.8	79.0	642	27	AA020934	AA020934 z664f12.r
C 15	15.8	79.0	305	28	AA120789	AA120789 zK90d11.r
C 16	15.8	79.0	433	30	AA248894	AA248894 JA00A091.
C 17	15.8	79.0	535	32	AA328931	AA328931 EST33605
C 18	15.8	79.0	450	33	AA417468	AA417468 m1/d01 Ye
C 19	15.8	79.0	453	33	AA433124	AA433124 JA00A201.
C 20	15.8	79.0	514	48	A1561906	A1561906 v183a09.x
C 21	15.8	79.0	337	50	AV047614	AV047614 AV047614
C 22	15.4	77.0	384	24	H72799	H72799 YU07d10.r1
C 23	15.4	77.0	334	27	AA050566	AA050566 m16a12.r
C 24	15.4	77.0	308	27	W91533	W91533 mF84c11.r1
C 25	15.4	77.0	554	34	AA524662	AA524662 nh35e08.s
C 26	15.4	77.0	310	39	AA831652	AA831652 oc82h10.s
C 27	15.4	77.0	380	46	A1443743	A1443743 sA45d06.Y
C 28	15.4	77.0	337	49	A1625843	A1625843 ty65d06.x
C 29	15.2	76.0	307	20	T05994	T05994 EST03883 Fe
C 30	15.2	76.0	494	20	T58174	T58174 YD26F04.r1
C 31	15.2	76.0	322	20	Z42752	Z42752 HSC08H081
C 32	15.2	76.0	322	21	F00595	F00595 HSB17G012 S
C 33	15.2	76.0	314	21	F13112	F13112 HSC3TH051 n
C 34	15.2	76.0	363	21	R09683	R09683 Y123h05.r1
C 35	15.2	76.0	342	21	T75528	T75528 Yd63d04.r1
C 36	15.2	76.0	349	21	T82265	T82265 Yd44g03.s1
C 37	15.2	76.0	381	22	R50603	R50603 YJ60g08.r1
C 38	15.2	76.0	417	22	R52980	R52980 YG85e10.s1
C 39	15.2	76.0	472	22	R53776	R53776 Y102e11.r1
C 40	15.2	76.0	426	22	R55507	R55507 YJ75g11.r1
C 41	15.2	76.0	445	22	R73800	R73800 YJ97g09.r1
C 42	15.2	76.0	456	22	R73843	R73843 Y155f11.r1
C 43	15.2	76.0	464	23	H44728	H44728 YP24f05.r1
C 44	15.2	76.0	444	23	H45554	H45554 Y072d05.r1
C 45	15.2	76.0	510	51	A1708406	A1708406 at13d04.x

ALIGNMENTS

RESULT 1
N44819/c N44819 475 bp mRNA EST 13-FEB-1996
LOCUS YY39a10.r1 Soares melanocyte 2Nbhm Homo sapiens
DEFINITION IMAGE:773594 5' similar to SW:P1A_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
RNA sequence.
ACCESSION N44819

NID 01185985
 VERSION N44819.1 GI:1185985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 LITERATURE 1 (bases 1 to 475)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Treviski, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The Washu-Merck EST Project
 JOURNAL Unpublished (1993)
 COMMENT On May 8, 1995 this sequence version replaced gi:800198.
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: T7
 High quality sequence stop: 402.
 Location/Qualifiers
 1..475
 /organism="Homo sapiens"
 /db_xref="GDB:3883236"
 /db_xref="taxon:9606"
 /clone_image="273594"
 /clone_lib="Soares melanocyte 2NBM"
 /sex="Male"
 /tissue="melanocyte"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCGCGAGTATTTTATTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M. Fatima Bonaldi. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 170 a 170 c 85 g 128 t 2 others
 ORIGIN

Query Match 100.0%; Score 20; DB 25; Length 475;
 Best Local Similarity 100.0%; Pred. No. 9.5;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGCTGTAATTCATATGCTG 20
 ||||||||||||||||||||
 Db 372 TGCTGTAATTCATATGCTG 353

RESULT 2
 AA298585 343 bp mRNA EST 18-APR-1997
 LOCUS EST114216 HSC172 cells II Homo sapiens cDNA 5' end similar to
 DEFINITION phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
 sequence.
 accession AA298585
 NID 01950938
 VERSION AA298585.1 GI:1950938
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE 1 (bases 1 to 343)
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, Wai, C., Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D., Fitzgerald, L.M., Fitch, W.M., Fitchman, J.L., Geoghagen, N.S., Goddek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr., Kelley, J.M., Kelley, J.C., Liu, L.-I., Marnaros, S.M., Merrick, J.M., Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudak, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bedarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, D.F., Ferrel, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meisner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M., Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.
 TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1392803.
 Other ESTs: TIGR168479
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlavage@tigr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/db/hgi/hgi.html>)
 Seq primer: M3 Reverse.
 Location/Qualifiers
 1..343
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):179595"
 /db_xref="taxon:9606"
 /clone_lib="HSC172 cells II"
 /cell_type="fibroblast"
 /cell_line="HSC172 (60PDL)"
 /dev_stage="fetal"
 /note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI;"

BASE COUNT 132 a 61 c 64 g 84 t 2 others
 ORIGIN

Query Match 100.0%; Score 20; DB 31; Length 343;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TGCTGTAATTCATATGCTG 20
 ||||||||||||||||||||
 Db 287 TGCTGTAATTCATATGCTG 268

RESULT 3
 A1632172 633 bp mRNA EST 26-APR-1999
 LOCUS ts85g07.x1 NCI-CGP6 GC6 Homo sapiens cDNA clone IMAGE:2238108 3'
 DEFINITION similar to SW.P11A-BOVIN P32871 PHOSPHATIDYLIINOSITOL 3-KINASE
 CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.
 accession A1632172
 NID 04683502
 VERSION A1632172.1 GI:4683502
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE *1 (bases 1 to 633)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 7, 1998 this sequence version replaced gi:3121411.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html

Seq primer: -400P from Gibco
 High quality sequence stop: 446.

FEATURES
 source

1..633
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="15q26.1"
 /clone="IMAGE:2238108"
 /clone_1lb="NCI-CGAP-GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT733-Pac (Pharmacia) with a modified
 polylinker; plasmid DNA from the normalized library
 NCI-CGAP-GC4 was prepared, and ss circles were made in
 vitro. Following HAP purification, this DNA was used as
 tracer in a subtractive hybridization reaction. The driver
 was PCR amplified cDNAs from a pool of 5,000 clones made
 from the same library (cloneids 1257096-1258631,
 1469064-1470983, and 1475592-1476743). Subtraction by
 Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 169 a 109 c 108 g 246 t 1 others
 ORIGIN

Query Match 100.0%; Score 20; DB 49; Length 633;

Best Local Similarity 100.0%; Pred. No. 8.8;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTTAATGCTG 20
 ||||||||||||||||
 223 TCGCTGAATCTTAATGCTG 239

RESULT 4

LOCUS N84360 136 bp mRNA EST 01-APR-1996
 DEFINITION KR7998F Human fetal heart, Lambda ZAP Express Homo sapiens CDNA
 clone KR7998 5' similar to PHOSPHATIDYLINOSITOL 3-KINASE, mRNA
 sequence.
 ACCESSION N84360
 NID N84360
 VERSION g1259985
 KEYWORDS GI:1259985
 EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 136)
 AUTHORS Liew, C.C.
 TITLE CDNAs from fetal heart (1996)
 JOURNAL Unpublished (1996)
 COMMENT On May 8, 1995 this sequence version replaced gi:800157.

Contact: Liew CC
 Department of Laboratory Medicine and Pathobiology
 University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: liewcc@utcc.utoronto.ca
 Seq primer: GAATTAACCTCACTTAAGG.
 FEATURES
 source

1..136
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="KR7998"
 /clone_1lb="Human fetal heart, Lambda ZAP Express"
 /lab_host="E. coli XL1-Blue"
 /note="Vector: Lambda ZAP Express; Site 1: EcoRI; Site 2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). CDNA was synthesized using a XhoI-Oligo dT
 adapter-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."
 BASE COUNT 52 a 29 c 26 g 29 t
 ORIGIN

Query Match 92.0%; Score 18.4; DB 25; Length 136;

Best Local Similarity 95.0%; Pred. No. 70;

Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTTAATGCTG 20
 ||||||||||||||||
 Db 124 TGCCTGAATCTTAATGCTG 105

RESULT 5

LOCUS AA508979 340 bp mRNA EST 08-JUL-1997
 DEFINITION MBACWIH03T3 Brugia malayi adult female CDNA (Saw95MLW-BMAC) Brugia
 malayi CDNA clone AFWIHO3 5', mRNA sequence.

ACCESSION AA508979
 NID 62246856
 VERSION AA508979.1 GI:2246856
 KEYWORDS EST.
 SOURCE Brugia malayi.
 ORGANISM Eukaryota; Metazoa; Nematoda; Secernentea; Spirurida; Spiruridae;
 Filarioidea; Onchocercidae; Brugia.

REFERENCE 1 (bases 1 to 340)
 AUTHORS Blaxter, M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and
 Jones, S.J.
 TITLE Genes expressed in adult female Brugia malayi
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395421.

Contact: Blaxter ML
 Institute of Cell, Animal and Population Biology
 University of Edinburgh
 Ashworth labs, King's Buildings, West Mains Road, Edinburgh, EH9
 3JF, UK.
 Tel: +44 131 650 6760
 Fax: +44 131 670 3450
 Email: mark.blaxter@ed.ac.uk
 The ABI trace of this sequence can be viewed at
<http://www.sanger.ac.uk/brugia/AFWC/MBACWIH03T3.html>
 Seq primer: T3

FEATURES
 source
 1..340
 /organism="Brugia malayi"
 /db_xref="taxon:6279"
 /map="891D05; 6; 6q22.2-6q23.2; 17q21"
 /clone="AFWCWIH03"
 /clone_1lb="Brugia malayi adult female CDNA

```

(SAM96MIM-Emaf)"
/dev-stage="adult"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from approximately 50 adult females
isolated from the peritoneal cavity of birds and
converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 5 x 10E6 independent recombinants
and the average insert size is ~900bp. The library was
constructed by Michelle Lizotte-Waniewski. The
library is available from Dr.S.A.Williams, email:
genome@mith.edu."
BASE COUNT      100 a      54 c      84 g      102 t
ORIGIN

Query Match      92.0%; Score 18.4; DB 34; Length 340;
Best Local Similarity 95.0%; Pred. No. 54;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 TGCTGTAATTCATGCTG 20
|||||
Db      101 TGCTGTAATTCATGCTG 120

RESULT 6
R10988      246 bp      mRNA      EST      11-APR-1995
LOCUS      yf38c09.r1 Soares fetal liver spleen INFLS Homo sapiens CDNA clone
DEFINITION IMAGE:129136 5', mRNA sequence.
ACCESSION  R10988
NID         9763723
VERSION     R10988.1 GI:763723
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1 (bases 1 to 246)
Hillier,L., Clark,N., Dubnue,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevaslis,E., Waterston,R., Williamson,A., Woldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
TITLE       The WashU-Merck EST Project
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert Size: 785
High quality sequence stops: 170 Source: IMAGE Consortium, LINT
This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 785 Std Error: 0.00
Seq primer: M13Rpl
High quality sequence stop: 170.
Location/Qualifiers
1. 246
/organism="Homo sapiens"
/db_xref="GDB:481297"
/db_xref="taxon:9606"
/clone="IMAGE:129136"
/clone_11b="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"

(SAM96MIM-Emaf)"
/dev-stage="adult"
/lab_host="XLI-Blue MRF"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from approximately 50 adult females
isolated from the peritoneal cavity of birds and
converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 5 x 10E6 independent recombinants
and the average insert size is ~900bp. The library was
constructed by Michelle Lizotte-Waniewski. The
library is available from Dr.S.A.Williams, email:
genome@mith.edu."
BASE COUNT      100 a      54 c      84 g      102 t
ORIGIN

Query Match      87.0%; Score 17.4; DB 21; Length 246;
Best Local Similarity 94.7%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY      1 TGCTGTAATTCATGCTG 19
|||||
Db      111 TGCTGTAATTCATGCTG 93

RESULT 7
C80053      580 bp      mRNA      EST      26-JUN-1998
LOCUS      C80053 Mouse 3.5-dpc blastocyst CDNA Mus musculus CDNA clone
DEFINITION J0075H01 3', mRNA sequence.
ACCESSION  C80053
NID         92520383
VERSION     C80053.1 GI:2520383
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 580)
Ko,M.S.H., Kitchen,J.R., Wang,X., Wang,X., Threat,T.A., Sun,T.,
Grubovack,M.J., Mason,S., Lim,M.K., Paonessa,P.D., Sauls,A.D. and
Doi,H.
Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst
(the ERATO/Doi Project at Wayne State University)
Unpublished (1997)
On Jan 24, 1995 this sequence version replaced gi:634372.
TITLE       Systematic analyses of genes expressed in 3.5-dpc mouse blastocyst
JOURNAL
COMMENT

Contact: Hirofumi Doi
Doi Biosymmetry Project, ERATO
Japan Science and Technology Corporation (JST)
WBG Marine East 12F, 2-6 Nakase, Mihama-ku, Chiba 261-71, Japan
Email: hdo@bioa.jst.go.jp.
Location/Qualifiers
1. 580
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="J0075H01"
/clone_11b="Mouse 3.5-dpc blastocyst CDNA"
/tissue_type="blastocyst"
/dev_stage="3.5-dpc"
BASE COUNT      168 a      134 c      134 g      137 t      7 others
ORIGIN

Query Match      84.0%; Score 16.8; DB 36; Length 580;
Best Local Similarity 90.0%; Pred. No. 2.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY      1 TGCTGTAATTCATGCTG 20
|||||
Db      68 TGCTGTAATTCATGCTG 49

RESULT 8
A102178.

```

LOCUS A1102178 208 bp mRNA EST 31-JAN-1999
 DEFINITION EST211467 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
 RRCGA22 3' end, mRNA sequence.
 ACCESSION A1102178
 NID 93706986
 VERSION A1102178.1 GI:3706986
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 208)
 Lee, N.H., Glodok, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
 Gene Index
 Unpublished (1998)
 JOURNAL On Aug 21, 1998 this sequence version replaced.
 COMMENT

CONTACT: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..208
 /organism="Rattus sp."
 /db_xref="ATCC (inhost):2024467"
 /db_xref="taxon:10118"
 /map="19p12-p13.1"
 /clone="RRCGA22"
 /clone_lib="Normalized rat brain, Bento Soares"
 /note="Organ: brain; Vector: pT7Tpac; Site_1: EcoRI;
 Site_2: NciI"
 BASE COUNT 57 a 53 c 34 g 64 t
 ORIGIN

Query Match 84.0%; Score 16.8; DB 43; Length 208;
 Best Local Similarity 90.0%; Pred. No. 3.2e+02;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 TGTCTAATCTATGCTG 20
 |||||
 Db 32 TGTCTGAATCTATGCTG 51

RESULT 9
 C83915/c 613 bp mRNA EST 28-APR-1999
 LOCUS C83915 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium
 discoideum cDNA clone SSA593, mRNA sequence.
 ACCESSION C83915
 NID 92706847
 VERSION C83915.1 GI:2706847
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum.
 Eukaryota; Dictyostelid; Dictyostelium.
 1 (bases 1 to 613)
 Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
 Yoshino, R., Mitra, B.N., Pl.M., Sato, T., Takemoto, K., Yasukawa, H.,
 Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
 The Dictyostelium developmental cDNA project: generation and
 analysis of expressed sequence tags from the first-finger stage of
 development
 JOURNAL DNA Res. 5 (6), 335-340 (1998)
 MEDLINE 99156227
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2153552.

CONTACT: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
 Email: d402huesakura.cc.tsukuba.ac.jp
 PROJECT - 'Dictyostelium discoideum cDNA project in Japan'
 POLYA-NO.
 Location/Qualifiers
 1..613
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SSA593"
 /clone_lib="Dictyostelium discoideum SS (H. Urushihara)"
 /dev_stage="sing"
 BASE COUNT 207 a 65 c 104 g 236 t 1 others
 ORIGIN

Query Match 80.0%; Score 16; DB 39; Length 613;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 4 TGTAAATCTATGCT 19
 |||||
 Db 211 TGTAAATCTATGCT 196

RESULT 10
 A0037810/c 580 bp mRNA EST 29-MAR-1999
 LOCUS A0037810 Dictyostelium discoideum SS (H. Urushihara) Dictyostelium
 discoideum cDNA clone SSE323, mRNA sequence.
 ACCESSION A0037810
 NID 93984563
 VERSION A0037810.1 GI:3984563
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum.
 ORGANISM Dictyostelium discoideum.
 Eukaryota; Dictyostelid; Dictyostelium.
 1 (bases 1 to 580)
 Morio, T., Urushihara, H., Saito, T., Ugawa, Y., Mizuno, H., Yoshida, M.,
 Yoshino, R., Mitra, B.N., Pl.M., Sato, T., Takemoto, K., Yasukawa, H.,
 Williams, J., Maeda, M., Takeuchi, I., Ochiai, H. and Tanaka, Y.
 The Dictyostelium developmental cDNA project: generation and
 analysis of expressed sequence tags from the first-finger stage of
 development
 JOURNAL DNA Res. 5 (6), 335-340 (1998)
 MEDLINE 99156227
 COMMENT On Jan 5, 1998 this sequence version replaced gi:2747316.

CONTACT: Hideko Urushihara
 Institute of Biological Sciences
 University of Tsukuba
 3-3-10 Ten-noda, Tsukuba, Ibaraki 305, Japan
 Email: d402huesakura.cc.tsukuba.ac.jp
 PROJECT - 'Dictyostelium discoideum cDNA project in Japan'.
 Location/Qualifiers
 1..580
 /organism="Dictyostelium discoideum"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="SSE323"
 /clone_lib="Dictyostelium discoideum SS (H. Urushihara)"
 /dev_stage="sing"
 BASE COUNT 217 a 53 c 86 g 224 t
 ORIGIN
 Query Match 80.0%; Score 16; DB 44; Length 580;
 Best Local Similarity 100.0%; Pred. No. 5.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 TGTAAATCTAATGCT 19
 Db 89 TGTAAATCTAATGCT 74

RESULT 11

LOCUS R52778 496 bp mRNA EST 18-MAY-1995
 DEFINITION Y996C01.r1 Soares infant brain INIB Homo sapiens cDNA clone
 IMAGE:41742.5', mRNA sequence.
 R52778
 NID 9814680
 VERSION R52778.1 GI:814680
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 496)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapel, B.,
 Chisoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, M.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
 Marids, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J.,
 Trevasaki, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu

Insert Size: 1524
 High quality sequence stops: 340 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1524 Std Error: 0.00
 Seq primer: M13RP1

High quality sequence stop: 340.

FEATURES

source

Location/Qualifiers
 1..496
 /organism="Homo sapiens"
 /db_xref="GDB:414283"
 /db_xref="taxon:9606"
 /clone="IMAGE:41742"
 /clone_1lb="Soares infant brain INIB"
 /sex="female"

/dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lambda BA; Site 1: Not
 I; Site 2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5';
 AACTGAGAGAAATTCGCGCCGAGAGAAATTTTTTTTTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the lambda BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 149 a 117 c 114 g 112 t 4 others

ORIGIN

Query Match 79.0%; Score 15.8; DB 22; Length 496;
 Best Local Similarity 89.5%; Pred. No. 6.7e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGTCTAATCTAATGCT 19
 Db 343 TGTCTAATCTAATGCT 325

RESULT 12

LOCUS R59012 500 bp mRNA EST 24-MAY-1995
 DEFINITION Y996C01.r1 Soares infant brain INIB Homo sapiens cDNA clone
 IMAGE:41331.5', mRNA sequence.
 R59012
 NID 9829707
 VERSION R59012.1 GI:829707
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 500)
 AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasaki, E., Waterston, R., Williamson, A., Wohlmann, P. and
 Wilson, R.

TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT On May 8, 1995 this sequence version replaced gi:800052.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu

Insert Size: 1241
 High quality sequence stops: 396 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1241 Std Error: 0.00
 Seq primer: M13RP1

High quality sequence stop: 396.

FEATURES

source

Location/Qualifiers
 1..500
 /organism="Homo sapiens"
 /db_xref="GDB:413872"
 /db_xref="taxon:9606"
 /clone="IMAGE:41331"
 /clone_1lb="Soares infant brain INIB"
 /sex="female"

/dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: lambda BA; Site 1: Not
 I; Site 2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer [5';
 AACTGAGAGAAATTCGCGCCGAGAGAAATTTTTTTTTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (Pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the lambda BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 137 a 145 c 127 g 86 t 5 others

ORIGIN

Query Match 79.0%; Score 15.8; DB 22; Length 500;
 Best Local Similarity 85.0%; Pred. No. 6.7e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 TGTCTAATCTAATGCT 20
 Db 441 TGTCTAATCTAATGCT 422

RESULT 13
 LOCUS N50767 214 bp mRNA EST 14-FEB-1996

DEFINITION Y990f02.r1 Soares multiple sclerosis 2NDHMSF Homo sapiens cDNA
clone IMAGE:280827 5', mRNA sequence.

ACCESSION N50767
NID 91191933
VERSION N50767.1 GI:1191933
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 214)
AUTHORS Hallier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The Mashu-Merck EST Project
Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:785418.

TITLE JOURNAL
MEDLINE
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: 7'
High quality sequence stop: 180.

FEATURES
Source
Location/Qualifiers
1..214
/organism="Homo sapiens"
/db_xref="GDB:389203"
/db_xref="taxon:9606"
/map="10"
/clone="IMAGE:280827"
/clone_lib="Soares_multiple_sclerosis_2NDHMSF"
/sex="male"
/tissue_type="multiple sclerosis lesions"
/dev_stage="Age 46"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker V-type; phagemid; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTACCAATCTGAAGTGGAGCGCGCCGATTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot - 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from 4 multiple sclerosis
lesions from one patient was kindly provided by Dr. Kevin
G. Becker (NINDS/NIH).

BASE COUNT 53 a 47 c 44 g 61 t 9 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 25; Length 214;
Best Local Similarity 89.5%; Pred. No. 8.7e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTTAATGCT 19
||||| ||||| |||||
Db 176 TGCCTGAATCTTAATGCT 194

RESULT 14
AA020934 642 bp mRNA EST 30-JAN-1997
LOCUS AA020934
DEFINITION z64f12.r1 Soares retina N2b4HR Homo sapiens cDNA clone
IMAGE:363791 5', mRNA sequence.

ACCESSION AA020934
NID 91484696
VERSION AA020934.1 GI:1484696
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 642)
AUTHORS Hallier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapell, B.,
Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hulman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Scheilenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
COMMENT On May 8, 1995 this sequence version replaced gi:799442.

TITLE JOURNAL
MEDLINE
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1229 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 412.

FEATURES
Source
Location/Qualifiers
1..642
/organism="Homo sapiens"
/db_xref="GDB:1280494"
/db_xref="taxon:9606"
/clone="IMAGE:363791"
/clone_lib="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCCGCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). The retina was obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Rodrick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M. Fatima Bonaldo.

BASE COUNT 185 a 119 c 148 g 183 t 7 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 27; Length 642;
Best Local Similarity 89.5%; Pred. No. 6.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCCTGAATCTTAATGCT 19
||||| ||||| |||||
Db 175 TGTGTGAATCTTAATGCT 193

RESULT 15
AA120789 305 bp mRNA EST 19-NOV-1996
LOCUS AA120789
DEFINITION zK90d11.r1 Soares_pregnant_uterus_NBHPH Homo sapiens cDNA clone

IMAGE:480101.5', mRNA sequence.

ACCESSION A1120789

NID 91677984

VERSION A1120789.1 GI:1677984

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 305)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Woldmann, P. and Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

COMMENT On May 8, 1995 this sequence version replaced gi:800961.

TITLE JOURNAL

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 255.

FEATURES

source

Location/Qualifiers

1..305

/organism="Homo sapiens"

/db_xref="GDB:380468"

/db_xref="taxon:9606"

/clone="IMAGE:490101"

/clone_lib="Soares_pregnant_uterus_NbHPV"

/sex="female"

/dev_stage="adult"

/lab_host="DH10B"

/note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I;

Site_2: Eco RI; 1st strand cDNA was primed with a Not I -

oligo(dT) primer [5'

AACGGAAGATTCCGGCCGCTTTTCTTTTCTTTTCTTTT 3']

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by M. Fatima Bonaldo."

BASE COUNT 78 a 60 c 62 g 96 t 9 others

ORIGIN

Query Match 79.0%; Score 15.8; DB 28; Length 305;

Best Local Similarity 89.5%; Pred. No. 7.8e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCCTGTAATCTAATGCT 19

|||||

DB 254 TCCTGTAATCTAATGCT 272

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:35 ; Search time 539.84 Seconds

(without alignments)
117.824 Million cell updates/sec

Title: US-09-325-095-22

Perfect score: 1 GATTTCATGAAACAAATGA 20

Sequence:

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*
1: gb_dal.*
2: gb_daz.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_pu.*
7: gb_pil.*
8: gb_pil2.*
9: gb_prl.*
10: gb_prl2.*
11: gb_prl3.*
12: gb_prl4.*
13: gb_prl5.*
14: gb_prl6.*
15: gb_prl7.*
16: gb_prl8.*
17: gb_prl9.*
18: gb_prl10.*
19: gb_prl11.*
20: gb_prl12.*
21: gb_prl13.*
22: gb_prl14.*
23: gb_prl15.*
24: gb_prl16.*
25: gb_prl17.*
26: gb_prl18.*
27: gb_prl19.*
28: gb_prl20.*
29: gb_prl21.*
30: gb_prl22.*
31: gb_prl23.*
32: gb_prl24.*
33: gb_prl25.*
34: gb_prl26.*
35: gb_prl27.*
36: gb_prl28.*
37: gb_prl29.*
38: gb_prl30.*
39: gb_prl31.*
40: gb_prl32.*
41: gb_prl33.*
42: gb_prl34.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	3207	3. BOVPHOS3KN	M93252 Bovine phop

2	20	100.0	3498	5	A37232
3	20	100.0	3424	10	HSPI3K
4	20	100.0	3207	10	HSU79143
5	17.4	87.0	12816	2	AF0292224
6	17.4	87.0	10670	5	A67161
7	17.4	87.0	81370	7	AB017071
8	17.4	87.0	8151	7	SCPR8GNA
9	17.4	87.0	7845	7	YSCDBF3A
10	17.4	87.0	41664	8	YSCDBF3A
11	17.4	87.0	122325	9	HS1044017
12	17.4	87.0	101076	9	HS37M17
13	17.4	87.0	23464	11	AF111168
14	17.4	87.0	28291	36	CEB5H9
15	17	85.0	114144	8	ATU78721
16	17	85.0	135214	11	AC004848
17	17	85.0	1048	36	AF014157
18	17	85.0	131971	42	AC004527
19	16.8	84.0	1634	1	SNAP
20	16.8	84.0	1289	1	STAGASP
21	16.8	84.0	3452	4	AF001076
22	16.8	84.0	1586	5	E03835
23	16.8	84.0	1558	5	E03836
24	16.8	84.0	2788	7	SCYLR037C
25	16.8	84.0	2366	7	SCYLR038C
26	16.8	84.0	104308	8	AC006228
27	16.8	84.0	156432	9	HS581F7
28	16.8	84.0	9339	10	HS1L25FL
29	16.8	84.0	170136	11	AC003082
30	16.8	84.0	191111	11	AC006080
31	16.8	84.0	156898	11	U82696
32	16.8	84.0	3389	17	AF001075
33	16.8	84.0	191857	34	CEY32B12
34	16.8	84.0	224525	34	CEY70C5
35	16.8	84.0	180549	35	AC007316
36	16.8	84.0	312403	35	AC007541
37	16.8	84.0	177453	35	AC007646
38	16.8	84.0	173838	35	AC007650
39	16.8	84.0	173087	35	AC007669
40	16.8	84.0	9540	36	CEY70C5A
41	16.8	84.0	1900	36	DROGBR
42	16.8	84.0	1896	36	DROGBRA
43	16.8	84.0	1853	36	DROGBRN
44	16.8	84.0	83340	37	AC004438
45	16.8	84.0	12029	37	AE001428

ALIGNMENTS

RESULT 1	BOVPHOS3KN	3207 bp	mRNA	19-AUG-1992
LOCUS	BOVPHOS3KN	3207 bp	mRNA	19-AUG-1992
DEFINITION	Bovine phosphatidylinositol 3-kinase	110 kDa subunit	complete	
ACCESSION	M93252			
NID	g163519			
VERSION	M93252.1	GI:163519		
KEYWORDS	phosphatidylinositol 3-kinase.			
SOURCE	Bos taurus	CDNA to mRNA.		
ORGANISM	Bos taurus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.			
AUTHORS	Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R., Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F., Hsu, J.J., Courtneidge, S.A., Parker, P.J., and Waterfield, M.D.			
TITLE	Phosphatidylinositol 3-kinase: Structure and expression of the 110 kD catalytic subunit			
JOURNAL	Cell 70, 419-429 (1992)			
MEDLINE	92354059			
FEATURES	Location/Qualifiers			
SOURCE	1. 3207			

CDS

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AA030698.1"
/db_xref="PID:9163520"
/db_xref="GI:163520"
/translation="MPRSSGELWGILHMPRIIVECLIPNGMIVTLECLREATLIT
IKHLEFKARKYPLHQLQDESSYIFVSVTOEAREBEFDETRRLCDLRLPOPLKVI
EPVGNREKILNREIGFALGMPVCEFDNVKDEPQDFRNILNVCKEAVDLRLDSPH
SRAMVYPPNVSSPELPHKTYNKLKDGQIIVAVIYSPNDKRTYLLKINDCVPE
OVIARIRKTRSMILSSSEQLKCYLEVOGKYLIVCGDEFLKRYLSQKYLISC
IMLGMPNIMLMAKESLSQLEPMDCTFMPYSRSTATPYMNGETSTRKSLVINSAL
RIKILCATYVANNRIDIKIYVGTIGEGELCNVMTQVPCSNPRNENLNDIY
IPDLPRARLCLSTICSVKGRGAKKEHCPPLMNGINLFDYDTLVSGKMLNLMVPVH
GLEDLNPFGVTSNPKETPCLEEDWESSVVKFPMDSVIEEHANVSREAGFSY
SHAGLSNRLADNELRENDKEDRLAICTRDPSETTEOKDFLMSHRHCVTIPILP
KILSVKNSRDEVAQMTCLVNDMPRIPEQAMELLDCNPDPMVRGFAVCLERYLT
DOKLSOYLQIOLVQVLYEQLDNLVRLKRLKALNORIGHEFWHLKSEMNKTVSQ
RFGILLESYCRACGMYLKLHNRQVAMEKLINLIDILKOEKDETOKQMKFLVQMR
RPEDNALOGFLSPNPAHOLGNLRECRISMSAKRPLMWNENDIMSELTPQNE
IIFKNGDDLRODMLTLOIRIMEINWOGDILRLPYGCCSTGCAGLIEVNSHT
IMQIOCKGKLGALQFNSHTLHQLKDKNKEITDAIDLTRSCAGTCAVFTLIGIG
DRHNSINWKDQGLFHDIFGHFLDKHKKRKYKREKREPVLTODFLVISGAOECT
KTREPEROEKCYKAYLAIROHANLFINLFMSMLGSGMPELOSPDIAVIRKTLALDK
TEOEALVEYFMQNDAHNGWTTKMDMIFTHIKQHALN"
BASE COUNT      1028 a      581 c      680 g      918 t
ORIGIN

```

```

Query Match      100.0%; Score 20; DB 3; Length 3207;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 1 GTAATTCATGAACAATGA 20
Db 3111 GTAATTCATGAACAATGA 3130

```

```

RESULT 2
LOCUS      A37232      3498 bp      DNA      PAT      05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
NID        A37294345
VERSION    A37232.1 GI:2294345
KEYWORDS
SOURCE     unidentified.
ORGANISM   unidentified.
REFERENCE  1 (bases 1 to 3498)
AUTHORS   Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE      EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL    Patent: WO 9403609-A 1 17-FEB-1994;
            IMP CANCER RES TECH (GB)
            Other publication JP 8503124T 960409.
FEATURES
SOURCE     Location/Qualifiers
            1..3498
            /organism="unidentified"
            /db_xref="taxon:32644"
            1..3207
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CA002284.1"
            /db_xref="PID:e304808"
            /db_xref="PID:92294346"
            /db_xref="GI:2294346"
            /translation="MPRSSGELWGILHMPRIIVECLIPNGMIVTLECLREATLIT
            IKHLEFKARKYPLHQLQDESSYIFVSVTOEAREBEFDETRRLCDLRLPOPLKVI

```

```

BASE COUNT      1147 a      626 c      722 g      1003 t
ORIGIN

```

```

Query Match      100.0%; Score 20; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Cy 1 GTAATTCATGAACAATGA 20
Db 3111 GTAATTCATGAACAATGA 3130

```

```

RESULT 3
LOCUS      HSPH13K      3424 bp      mRNA      PRI      24-AUG-1995
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION  Z29090
NID        9472990
VERSION    Z29090.1 GI:472990
KEYWORDS   phosphatidylinositol 3-kinase.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominoidea; Homo.

```

```

REFERENCE  1 (bases 1 to 1068)
AUTHORS   Volinia,S., Hilsen,I., Ormondtroyd,E., Nizetic,D., Antonacci,R.,
            Rocchi,M. and Waterfield,M.D.
TITLE      Molecular cloning, cDNA sequence, and chromosomal localization of
            the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL    Genomics 24 (3), 472-477 (1994)
MEDLINE    95229146
REFERENCE  2 (bases 1 to 3424)
AUTHORS   Volinia,S.
TITLE      Direct Submission
JOURNAL    Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
            Institute for Cancer, Research, 91 Riding House Street, London, W1P
            8BT, UK

```

```

FEATURES
SOURCE     Location/Qualifiers
            1..3424
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3q26.3"
            /cell_line="K562"
            13..3219
            /codon_start=1
            /product="phosphatidylinositol 3-kinase"
            /protein_id="CA082333.1"
            /db_xref="PID:9472991"
            /db_xref="GI:472991"
            /db_xref="SWISS-PROT:P42336"
            /translation="MPRSSGELWGILHMPRIIVECLIPNGMIVTLECLREATLIT
            IKHLEFKARKYPLHQLQDESSYIFVSVTOEAREBEFDETRRLCDLRLPOPLKVI
            EPVGNREKILNREIGFALGMPVCEFDNVKDEPQDFRNILNVCKEAVDLRLDSPH
            SRAMVYPPNVSSPELPHKTYNKLKDGQIIVAVIYSPNDKRTYLLKINDCVPE
            OVIARIRKTRSMILSSSEQLKCYLEVOGKYLIVCGDEFLKRYLSQKYLISC

```


INIGRPNLAKMAKESLYSOLPMDCTMPSYRSRISTATPYMNGESTSKSLVATNRL
 RIKITLCATYNNLRNDIDKTYVRGTYHGESEPCDNYNTQRPVSNRNMENLNTDYL
 IPDLPRARICLVSICSVKGRKGAKEEHCPLAMGNI.NLFDYTDITLVSGKMLNMPVPH
 GLBIDLNPICVTSNPKETPCLEEDPMSVVKPDMASVIEHANMSYREGEFSY
 SHAGLSNRLARNDRENDKEQLKAISTROPLESTIOEDFLSHRHHTCYTITIELP
 KLLISVWNSRDEVAQMYCLVKMPPIKPPQAMELLDCNYPDPVNRGFAVCLEKYL
 DDKLSOYLIOIYVLYKYEOLYLDNLVRELLKALTNORIGHEFFMHLKSEHNNTVSO
 RPDGLISYCRACGMYLKHLNROYEAMEKILNLTDLIOEKREKDETQVOKFLEVQNR
 RPFMDALQGLISPLNPAHOLGNRLKECRIMSSAKRPMLNMENPDMISELLEFQNE
 IIFKNGDDLKQDM.LTIQIRIMENIMONOGDLMLPYGLSICDGLLEVYNSHT
 IMOIOCKGLKGLAOLFNSHTLHOWLKDKNKGSEIIDAIDLETRSCAGCYATFIIIGIG
 DRHNSNTMVKDDGOLFPHIDFGHFLDHKKKFGYKREVRPEVLQDFLIIVISKGQECT
 KTRFERFQEMCYKAYLAIRQHANLFTNLSMMLSGMPQLQSFDDIAYIRKTLALDK
 TEQALTEYFKQMDAHHGCTTKMDWIFHTIKOHALN"

BASE COUNT 1134 a 618 c 709 g 963 t

ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 3424;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
 ||||||||||||||||
 Db 3123 GTATTTCATGAACAATGA 3142

RESULT 4
 HSU79143 3207 bp mRNA PRI 02-JAN-1997
 LOCUS Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA,
 DEFINITION complete cds.
 ACCESSION U79143
 MID G1763625
 VERSION U79143.1 GI:1763625
 KEYWORDS human.
 SOURCE Homo sapiens
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 3207)
 AUTHORS Stirdivant,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
 Oliff,A. and Heimbrock,D.C.
 TITLE Catalytic Activity of the p110-alpha Subunit of Human
 Phosphoinositide 3'-Hydroxykinase is Required for Signal
 Transduction
 JOURNAL Bioorg. Med. Chem. (1996) In press
 REFERENCE 2 (bases 1 to 3207)
 AUTHORS Stirdivant,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
 Oliff,A. and Heimbrock,D.C.
 TITLE Direct Submission
 JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
 Sumneytown Pike, West Point, PA 19486, USA
 FEATURES
 source location/Qualifiers
 1..3207
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 1..3207
 /function="phospholipid kinase"
 /product="phosphoinositide 3'-hydroxykinase p110-alpha
 subunit"
 /protein_id="AAB39753.1"
 /db_xref="PID:G1763625"
 /db_xref="GI:1763625"
 /translation="MPRRSSGELMGTHIMPRPLIVCEILPNGMIVTLECREATILT
 IKHLEKARKYPLHQLQDSSSTIFSYVEAREREFDETRRLCDRLRQPLKAYI
 EPVGNREKILNREIGFALIGMPCEVMDKPEVODFRRLTLNVCKRAVDRLDINSPE
 SRAYVYPPNVESSELPKHLIYNKLDGQIIVVIVVSPNNNDKQKTYLKINHCVBE
 QVIAEAIKRTKRLSSEQLKCLVLEVOGYILKVGCDYEFLEKYPISQYKTRSC
 INIGRPNLMAKESLYSOLPMDCTMPSYRSRISTATPYMNGESTSKSLVATNRL
 RIKITLCATYNNLRNDIDKTYVRGTYHGESEPCDNYNTQRPVSNRNMENLNTDYL
 IPDLPRARICLVSICSVKGRKGAKEEHCPLAMGNI.NLFDYTDITLVSGKMLNMPVPH

GLBIDLNPICVTSNPKETPCLEEDPMSVVKPDMASVIEHANMSYREGEFSY
 SHAGLSNRLARNDRENDKEQLKAISTROPLESTIOEDFLSHRHHTCYTITIELP
 KLLISVWNSRDEVAQMYCLVKMPPIKPPQAMELLDCNYPDPVNRGFAVCLEKYL
 DDKLSOYLIOIYVLYKYEOLYLDNLVRELLKALTNORIGHEFFMHLKSEHNNTVSO
 RPDGLISYCRACGMYLKHLNROYEAMEKILNLTDLIOEKREKDETQVOKFLEVQNR
 RPFMDALQGLISPLNPAHOLGNRLKECRIMSSAKRPMLNMENPDMISELLEFQNE
 IIFKNGDDLKQDM.LTIQIRIMENIMONOGDLMLPYGLSICDGLLEVYNSHT
 IMOIOCKGLKGLAOLFNSHTLHOWLKDKNKGSEIIDAIDLETRSCAGCYATFIIIGIG
 DRHNSNTMVKDDGOLFPHIDFGHFLDHKKKFGYKREVRPEVLQDFLIIVISKGQECT
 KTRFERFQEMCYKAYLAIRQHANLFTNLSMMLSGMPQLQSFDDIAYIRKTLALDK
 TEQALTEYFKQMDAHHGCTTKMDWIFHTIKOHALN"

BASE COUNT 1043 a 586 c 670 g 908 t

ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
 ||||||||||||||||
 Db 3111 GTATTTCATGAACAATGA 3130

RESULT 5
 AF029224 12816 bp DNA BCT 23-APR-1999
 LOCUS Staphylococcus carnosus nlr and nar operons, complete sequences.
 DEFINITION AF029224 AF029225
 ACCESSION G4596720
 MID AF029224.2 GI:4596720
 VERSION AF029224.2
 KEYWORDS Staphylococcus carnosus.
 SOURCE Staphylococcus carnosus
 ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group;
 Bacillus/Staphylococcus group; Staphylococcus.
 REFERENCE 1 (bases 6052 to 12816)
 AUTHORS Pantel,I., Lindgren,P.E., Neubauer,H. and Goetz,F.
 TITLE Identification and characterization of the Staphylococcus carnosus
 nitrate reductase operon
 JOURNAL Mol. Gen. Genet. 259 (1), 105-114 (1998)
 MEDLINE 98409266
 REFERENCE 2 (bases 1 to 6064)
 AUTHORS Neubauer,H., Pantel,I. and Goetz,F.
 TITLE Molecular characterization of the nitrite-reducing system of
 staphylococcus carnosus
 JOURNAL J. Bacteriol. 181 (5), 1481-1488 (1999)
 MEDLINE 99175446
 REFERENCE 3 (bases 1 to 12816)
 AUTHORS Neubauer,H., Pantel,I. and Goetz,F.
 TITLE Direct Submission
 JOURNAL Submitted (26-SEP-1997) Mikrobielle Genetik, Universitaet
 Tuebingen, Waldhauser Str. 70/8, Tuebingen 72076, Germany
 COMMENT On Apr 21, 1999 this sequence version replaced gi:3929521
 gi:4433635.
 FEATURES
 source location/Qualifiers
 1..12816
 /organism="Staphylococcus carnosus"
 /strain="TM300"
 /db_xref="taxon:1281"
 226..1056
 /gene="nirc"
 226..6032
 /gene="nir operon"
 226..1056
 /gene="nirc"
 /function="putative nitrite transporter"
 /codon_start=1
 /transl_table=11
 /product="nirc"
 /protein_id="AAB00822.1"
 /db_xref="PID:G4433636"
 /db_xref="GI:4433636"

```

/translation="VWMOHKTEVDYTAARETMNSVSSOMKSYMLSKTPVRYFLKA
MMSGFLSIVAFYMAIAIKQHGINGEBTINLGAIAFSAIALVILYTLSELTSPFMX
MTAGAVIKSVPIGKIWTWTFVCFIGNIVGGFVLGULAKTSVTPPEMVAIKIYDK
TTLTGWDGILVAIFANFPIKIDIVYSIQFKGLSKAFPIAGVVFVFMGEHYVN
ANLFSGMEFYNDALSMIDVLKNIMAFVAGNIGGIFVGLLYAFENGKRNHYIDKOK
"
gene      1184..1906
CDS       /gene="nir"
          1184..1906
          /gene="nir"
          /function="Involved in nitrite reduction"
          /codon_start=1
          /transl_table=11
          /product="Nir"
          /protein_id="AAD20823.1"
          /db_xref="PID:94433637"
          /db_xref="GI:4433637"
/translation="MSQIIYVHKMRGRKONOLMELEHVGKOVNGEEDVAFIESDE
LSLPVIRKYLAGEHEFTIVPLLFSGHYTELPTIMKEMQMPGSIHQIROP
CYHPALTNWIKRIDAKRYONNSNAIYIGSGSPFLTEPDELMLKRCOTECPIY
TMETGDOFEKLEPKMDNYHILLIPVFYDFGLVKNIKKINDIGNSDMTIAHA
LNFEPIDEMLAARLNEKES"
gene      1907..2362
CDS       /gene="sira"
          1907..2362
          /gene="sira"
          /note="similar to N-terminus of Escherichia coli CysG."
          /codon_start=1
          /transl_table=11
          /product="Sira"
          /protein_id="AAD20824.1"
          /db_xref="PID:94433638"
          /db_xref="GI:4433638"
/translation="WPIOLMLKMRVALIGGRIQFRKFKOLARADCEVTISKAP
WPEFHSYPIGKIITKDYDKEDIDADVVIATDSOEINDIKODTTPQOLVNHGTD
KSDSDFNMIEDFEDLAISVNSGDKYKVAISKAIKRYIQEYGR"
gene      2366..4771
CDS       /gene="nirB"
          2366..4771
          /gene="nirB"
          /note="similar to Escherichia coli NADH-dependent nitrite
          reductase large subunit NirB."
          /codon_start=1
          /transl_table=11
          /product="NirB"
          /protein_id="AAD20825.1"
          /db_xref="PID:94433639"
          /db_xref="GI:4433639"
/translation="MSKORLVWIGMGAVRTIEEIIIDRDPKFDITIIKRPYNY
RIMLSNIIOKMTVEETIMNPYSWEENGIALITGEKANRVNREKHIVETEGREPIY
DIGITATGSKAFVLPFGHELPSVIGWRTIEDTEKIEINODKKEATYVIGGLLEEC
ARGLDQGNVTVVLAELMEMOLDROAGDMLREDLEQCKFELGANFKETIENOD
VEAIRLADGRDLPADLVVAVAGIKPETOAIIDSGIDVERGIYVDPMRTSDEIYAVG
BCAQNKCYYGGLVLEQGVLAHLTGREGYQSTTFSLVSGCDLTSACIOQ
EDDEVEGIEVNSHDHYKKIPLKEERYVAVLYGDTDGTFRFNMKREPIEDYTL
VSLQKGEAVASTVADMDSETIGGNCVSGKIVSAIVEDGLTSAEYKATAGN
SGCKRPOIGELLEHTLGGAFAASKPAGICDTELTROIVQITAKGIKSKETRYH
IDMKNGGCPKPCRAINITYLNNVPHHEDEKASFPANERHANIONGTSVIOQR
GGVITPDQILRGEVAKRYDVLVTVGSQRIGLISIEDLPKWKDLGMSASAVA
KTRISVSKSCVCKEFCFGTQYTRIGRIEETFEYIDTPHKFMGSGCPRSCVSG
KDFGVIGENGFOIYIGNGGTETVKGGLTVEEEDVIOCGAYMOYRETGYAE
RTAPWLERMGFEFNVEYVLDPERKALVDRIIDAKRAVAEPEMSIIVDQDKIKFY
EKV"
gene      4774..5088
CDS       /gene="nirD"
          4774..5088
          /gene="nirD"
          /note="similar to Escherichia coli NADH-dependent nitrite
          reductase small subunit NirD."
          /codon_start=1
          /transl_table=11
          /product="NirD"
          /protein_id="AAD20826.1"

/db_xref="PID:94433640"
/db_xref="GI:4433640"
/translation="WRALEKRVAALEDLPILGTKVLRDKEIGLEITJEDGRJFAN
NVCPHKOPLSEEGVSEHSVCPPLHQDIDLETGLVQEPDTCVETEEVEIIDGVYL
CL"
gene      5079..6032
CDS       /gene="sirr"
          5079..6032
          /gene="sirr"
          /note="similar to C-terminus of Escherichia coli CysG."
          /codon_start=1
          /transl_table=11
          /product="Sirr"
          /protein_id="AAD20827.1"
          /db_xref="PID:94433641"
          /db_xref="GI:4433641"
/translation="MSVSPGTGHVILIGAPGNPMLTRKAEHLIRIADIIYDRLVN
FILOLAPETIETIDVKKPYSRHIQOSEINVELDAAYKHNNVYRIKGGDAIFGRV
EVDLIKNTPCETIVRGITTAASAANLIGLGRNSTSVTLTSGFRKSEQIED
IASLIDGSLAIYKIRLHTITTOIKQVKQDYPICAVFVSATQCVITGHSIDE
KRVQANAEPAEDAGILLIGDALEYADENTHIQLEETSKYFGECEIVTGREALIKAE
IYDAGCMCIDPNSDDHYHQSDLLIYRYMVDNNTIIVKA"
gene      6325..12810
CDS       /gene="nar operon"
          6325..9999
          /gene="narG"
          6325..9999
          /gene="narG"
          /note="similar to Escherichia coli nitrate reductases NRA
          and NRZ alpha subunits."
          /codon_start=1
          /transl_table=11
          /product="NarG"
          /protein_id="AAC82542.1"
          /db_xref="PID:93929522"
          /db_xref="GI:3929522"
/translation="MAKGMNFERKTEKENGMSVLTDKSNEMKMYREMSHDKYVR
TTHGYNCGSCSKAYFYKNGYITWENQOTDTPSCGPMPEPEPFCRGASFSYTES
PLRVRYPIYRKGLMELWTLALEHEGNNAIWAASIVDEPERAKIYNKARGGIMRYS
WKDVLQILAQOLLTYTIKKDGPRLAGFTPIPAMSLVASCARFINLARGMLSFYWM
YADLPASPOJMGEDTDVPESSDWNANVILIMGSNVLRTTPDAHFTVEYRGYKY
VSAVDYENYKFPADNMLAPNPGDAAVQAOMTAVILOKRYVDESEFVYAKOYTD
MPLIMDEHHEGSKAGRPILASDLGSENEKRPYIVDSEGITTPNMGORAE
EGKONMLKLENEDSKYDPAISVABGEDHIDIMKFPYFDNAGNITFERPIASREYTLA
DKRVKATVTDLMTSQTGIRFNHELEAKFDEAETPTPAMOEAAVAGVQDVIYTOV
ATERAONAIDTGRSMIIMGAGINWMSDSYIYAVNLWMLCCOQNGSGMAHYVG
QEKCRPIEGMSTIAFKDMQGPRLQMSYFTTQOMKYEEGVKLSPLAGLEK
HKHPADYVNTAARGLMPSYPOFTNSLRFGEDEKAEKEGFEENEYVLRKAVSVSRET
KEAVDEPRLRTNNHKSLEIYWSNLSLSSAGOEYFMKMLGTSGSGLLAPEDEKPEE
MIMWDVQGRKIDLVALDFRNTSTPLKADLYLPAAIYTERKHDISTDMHPVHPNPA
IDPLMESRSDMDIRKSQIVSEMAKYMPETVYDVVTSPLAHSKOEIATPYGVKD
WTKEVEEIPKRYMGFAVVDRTYLLYDKYVTAGPLLETAKIAGHGSFVPRKYOE
AKDIDVWSDCTIIDGRRIDTARADAVYINISANGELSOASYDLEKTMPLT
DISDSRSEKRTIFNNITSOPREVIPTAVESGSKNRYSPETNIEBLVFPRLTGR
OSYVLIHFQOFSQSLPYVKPTLPVAVFGRKQVIGGOONLTLRYLTPGKNHS
TYQNOHMLTIFRGGPPTWIAKEDAAHDHNDNLPVYNNRNGVYTAATYSHPMG
TMEFHADKHTIQTPGSEITGLRGSHAPRIHLPTQLVGAITQIYSTNYTPIG
NQORVYAVAKREKVEDLED"
gene      9989..11566
CDS       /gene="narH"
          9989..11566
          /gene="narH"
          /note="similar to Escherichia coli nitrate reductases NRA
          and NRZ beta subunits."
          /codon_start=1
          /transl_table=11
          /product="NarH"
          /protein_id="AAC82543.1"
          /db_xref="PID:93929523"
          /db_xref="GI:3929523"
/translation="MKTKAQYAVNLNDKICGTCGTCVTKNTWTRNPAEYIWFNNV
ETKPGVGYPRKMEQDEHYKGGMTLNKKGIDLKSGMSKIAFGKITVYNDPMVTKYX
YEPMTYVNEHTITAKASKHSVPATAHSAVISGDKIDLMLGWPMWEDDLAGHVTGPEDN

```

IORIEEKIOEDFOOTEMMYLPRICEHCLINPSCVASCPSGAMKREDEGCIYVLPODACR
GMRICMTGCPYKATYFMNKNKAKCTFCPPRVAPAGMPTVSCSTCTGMRATIGLTLID
ADRVHEVASCENEDDLYEKOCEFLNLPYDEVLIOAKDGIITMDMIEAONSPLYKLA
IEKAPLPLREYRTPLPMVYCPPLSPIMNFEKDSIKNDALFPAIEERLIEYL
ASLITAGDTKSVKALORAMAMRSYMAOYKGFEDIDRLARLITLEROOTDMRLLA
IAKEDFVITPSHKEEYMDTYAAGSGYTFGQNMGTGSGSGCEGCAVANKPGQ
TGOAVNESFPGGIFRD"

Query Match 87.0% Score 17.4; DB 2; Length 12816;
Best Local Similarity 94.7% Pred. No. 2.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATG 19
DB 8469 GTATTTCATGAACAATG 8487

RESULT 6
LOCUS A67161 10670 bp DNA PAT 05-MAY-1999
DEFINITION Sequence 1 from Patent EP0805205.
ACCESSION A67161
NID 94756148
VERSION A67161.1 GI:4756148
KEYWORDS
SOURCE Staphylococcus carnosus.
ORGANISM Staphylococcus carnosus
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/staphylococcus group; Staphylococcus.
REFERENCE 1 (bases 1 to 10670)
AUTHORS Fast,B., Galer,W., Goetz,F., Lindgren,P., Neubauer,H. and Pantel,I.
TITLE Nitrate reduction system of Staphylococcus carnosus
JOURNAL Patent: EP 0805205-A 05-NOV-1997;
NESTLE SA (CH)

FEATURES
source Location/Qualifiers
1..10670
/organism="Staphylococcus carnosus"
/db_xref="taxon:1281"
-35_signal
-10_signal
116..122
167..171
181..2586
CDS
/function="NIRB GENE"
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB42152.1"
/db_xref="PID:e1429587"
/db_xref="PID:g4756149"
/db_xref="GI:4756149"

/translation="MSKQRLVMIGNAGVTTIEIIDRPDKEDITIGKEPYNTN
RIMLSNLIKMTVEETIMNPYSWEENGILALITGEKATRYNREKHAIETEGEIPY
DICIFATGSKAFVPIPGHELPSYIGMTIEDTIRMIEIADKKKATVIGGLGLEC
ARGLDGGMNVTYVLAEMEMOLDROAGDMLREDLEKMGKFEIANTKEIIGND
VEARLADGRDLPADLYVMAGVIRPTQEDALDSIDYBERGIVYDNPARTSPETIYAG
ECAQHCNKVIGLVAPLIEQGVADHDLHGETEGYQSTITSKVSQSDYISAGQIQ
EDDEGEVEVINSNDHYKIKLKEERYGVALVDGDDGTRFNMAKRRPIDYTL
VSLQKVEEAVSTADMDSEITICGNGVSKGISALYEDGLTSALEYAKRAGN
SCGCKRPOIGLELLETGAGFAVASKPAGICCTELFROIVTOIRAGLAKTSKVBH
LDKNKGGCPCRCRAINYLYMYPHEHEDEKASRPANERYHANIONDGTSTVYPOK
GGYTNPDQILRLGEVAKKYDPLVYKTSQRTIGLXGSKEDLPYWKDQKRSATA
KTRRSVSCVKEKRCFTQYTRRLGRLLETIFY IDTPHKFMGVSCPPSCVSEGV
KDFGVIGVENFOIYIGNGTEVTGKOFITVETEDVDIOLCGAYMYRETYIYAE
RTAPWLERMGEENKYEYVLDPEROKALYDRINDAKRAVAEPMSEIVNDQDKIFKY
EKV"

CDS

2589..2903
/function="NIRD GENE"
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB42154.1"
/db_xref="PID:e1429590"
/db_xref="PID:g4756150"

terminator
-35_signal
-10_signal
4037..4042
4057..4062
4128..4134
4140..7814
CDS

CDS

CDS

CDS

/db_xref="GI:4756150"
/translation="MRAIEKRVALLDLPLPLGTLYLVNDKEIGLFTLEDGRIFAVN
MVCPHKGPSSEGVTSBHSVFCPLHDDQIDLETGLVDEPPTGCVETYEVEITDDVYL
CL"
3878..3878
4037..4042
4057..4062
4128..4134
4140..7814
/function="NARG GENE"
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB42156.1"
/db_xref="PID:e1429593"
/db_xref="PID:g4756151"
/db_xref="GI:4756151"
/translation="MAKFGMFFKPTKFNQNSVLTDKSEKEMKREKSHKYYR
TTHGNCCTGSCSMKVFVNGVITWENQOTIDPSCGPMPEPEPRGCFRGASFWEYS
PLRVRYPMRKLWELMTTALIEHEGNNIAAMASIVBDEKAKIKYKARGGHMRVS
WKDYLQILAAQLITTKKDGPDRIAGTPIPAMSMLSYASGARFINMGMLSEFYM
YADLPASPOIMGEOTDVPSESDWNNANYIIMGNSVPLRTTPAHMTEVRYGTGY
VSAVDYAEENYKFMADMLAPPGDAAVAAQMTHTIIOKYVYDSEMFVYAAQYD
MPILMFDEHGGYKACRPLASDLGHESENKRPYIVDNESEITIPYNTQORNE
EGKQWNLKLENEDESKVDPALSVAGEHDIDIMKFPYDNGNIGFERPIASREVTLA
DGKRVKATVYDMLTQYGIKRFNHELEAKGFDAETPYTPAMQAVAGVQDVTQV
AIEFAQNAIDTIGSRSMIIMGAGIHNMSDITIVAVNLVLCGQCVNGGGMAYTG
OEKCRPIEGMETIEMKQMOGPRLONGTSFVYTLTQMWKEEGSVKLASPLAGEIK
KHHPADYNAIRALGMLPSYPOFNTSLRGEDEKKEAGETNEEVLRAVSYVSRET
KFAVEDDLRTNHRKSLFIRNSNLSASAKOEFYFMHMLCTSSGLLAEPHEDEKPEE
MIMWDDVQGIIDLIVLDFRSTPLADLYLVPLATVYERHDSLSTDMHPVHPFND
IDPLMESRSDMDIFKSLQYVSEMAKYMPTQYDVVYSPLAHDSKDEIAPYKQVDE
WTKGEVAPLPGKTPGPAVVDROYTLIDRYTAGPLETAKLAGHVSFEVKQYEE
AKOIVGTMSDGTIKGRPRIDTARADVYVNLISATNGELSOASYKDLBETGCMPLT
DISDSRAEKTEEMNITSOPREVIPTAVFSGSNKRNRSPTTINIRIELYFRTLGR
QSYLDHEITQOEGESLPIYKPTLPYVFGKQDQYVGGQDNVLRILTLPKKNHIS
TYQDNQMLTLFRGGFTWLAKEDAEHHDHNDWLEVINNGVYTAAYVSHRPGG
TMFYHAQDNKHIOPTGSEITGLRGSSNAPRRLHKLPTQLGVYAQISTSFNYPGIG
NORDVYVAVRMEVDWLED"
7804..9381
/function="NARH GENE"
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB42158.1"
/db_xref="PID:e1429596"
/db_xref="PID:g4756152"
/db_xref="GI:4756152"
/translation="MKIRAOVAVYVNLNDKICIGHCVCYTKNTNRPAGYIMFNV
ETKRGVGYPRKMEDEOEHYKGGMTLNKKGKIDLKAGSSMSKJAFGRIFYNPMPIKXY
YEPWTYVNEHLITPAKASHSPVATASVISGDKIDLHMGPMWEDLAGHVTGCEDN
IORIEEKIOEDFOOTEMMYLPRICEHCLINPSCVASCPSGAMKREDEGCIYVLPODACR
GMRICMTGCPYKATYFMNKNKAKCTFCPPRVAPAGMPTVSCSTCTGMRATIGLTLID
ADRVHEVASCENEDDLYEKOCEFLNLPYDEVLIOAKDGIITMDMIEAONSPLYKLA
IEKAPLPLREYRTPLPMVYCPPLSPIMNFEKDSIKNDALFPAIEERLIEYL
ASLITAGDTKSVKALORAMAMRSYMAOYKGFEDIDRLARLITLEROOTDMRLLA
IAKEDFVITPSHKEEYMDTYAAGSGYTFGQNMGTGSGSGCEGCAVANKPGQ
TGOAVNESFPGGIFRD"
9374..9949
/function="NARI GENE"
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB42160.1"
/db_xref="PID:e1429599"
/db_xref="PID:g4756153"
/db_xref="GI:4756153"
/translation="MINIDNFKYQESIGYFAQQLNPEKLTFFHPKTFNDFITNNHRA
YAEVAVAYRDIQOSTLTETIKOLYTDYDFNKQAPLVMTYNNKFTQKRGOMLAKLYL
YEMFGLMEVNEHLSDYLPIMLEFYLAQWOMDDBAEGNIEFTIIVIDGSAVMQQLR
EIDSPVAPLIKGLRLETIKLQPEEVSHPHA"
9942..10625
/function="NARI GENE"

```

/note="unnamed protein product"
/codon_start=1
/translation_table=11
/protein_id="CAB42162.1"
/db_xref="PID:e1429602"
/db_xref="PID:e4756154"
/db_xref="GI:4756154"
/translation="MLNQLWVTPYLCVAIFIGHIANRIRDOESWAKSEFEIEK
RLKWSGLFHLGIIPIVAGHFGVLVPAALBEGMGNHIIYHIGAVYISLFGFVLI
GMILLYRVSINKIRLSASDMVNFELVICGLSTVTNNAISEPFYRQIIS
IMFRHLFTFNPNASLMTNVPWSPFKLHIFGLFSIFALMPFTLVHVMVSPVLSANRSYI
IYRKNRR"
BASE COUNT      3550 a      1771 c      2304 g      3045 t
ORIGIN

Query Match      87.0%; Score 17.4; DB 5; Length 10670;
Best Local Similarity 94.7%; Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GATTTCATGAACAATG 19
        |||||
Db      6284 GATTTCATGAACAATG 6302

RESULT 7
AB017071      81370 bp      DNA      PLN      06-FEB-1999
LOCUS      Arabidopsis thaliana genomic DNA, chromosome 3, pl clone: MSJ11,
DEFINITION      complete sequence.
ACCESSION      AB017071
VERSION      93510347
KEYWORDS      AB017071.1 GI:3510347
SOURCE      HTG.
            Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
            clone:MSJ11.
ORGANISM      Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsia.
            1 (sites)
REFERENCE      Nakamura,Y.
AUTHORS      Structural Analysis of Arabidopsis thaliana Chromosome 3. I
TITLE      Unpublished (1998)
JOURNAL      2 (bases 1 to 81370)
REFERENCE      Nakamura,Y.
AUTHORS      Direct Submission
TITLE      Submitted (26-AUG-1998) to the DDBJ/EMBL/GenBank databases.
JOURNAL      Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
            Gene Structure 2; 1532-3, Yana, Kisarazu, Chiba 292, Japan
            (E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
            Fax:+81-438-52-3934)
FEATURES
    source
        1. 81370
            /organism="Arabidopsis thaliana"
            /strain="Columbia"
            /db_xref="taxon:3702"
            /chromosome="3"
            /clone="MSJ11"
            /clone_lib="Mitsui P1"
BASE COUNT      25619 a      14925 c      14478 g      26348 t
ORIGIN

Query Match      87.0%; Score 17.4; DB 7; Length 81370;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 TATTTCATGAACAATG 20
        |||||
Db      14778 TATTTCATGAACAATG 14796
        |||||
TY1

```

```

RESULT 8
SCPP8GNA      8151 bp      DNA      PLN      17-FEB-1997
LOCUS      S.cerevisiae PRP8 gene, complete CDS.
DEFINITION      224732
ACCESSION      9395156
VERSION      224732.1 GI:395156
KEYWORDS      PRP8 gene; splicing factor.
SOURCE      Baker's yeast.
ORGANISM      Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
            Saccharomycetaceae; Saccharomyces.
            1 (bases 1 to 8151)
REFERENCE      Hodges,P.E., Jackson,S.P., Brown,J.D. and Beggs,J.D.
            Extraordinary sequence conservation of the PRP8 splicing factor
            yeast 11 (4), 337-342 (1995)
JOURNAL      95304847
MEDLINE
REFERENCE
AUTHORS      2 (bases 1 to 8151)
TITLE      Jackson,S.P., Lossky,M. and Beggs,J.D.
            Cloning of the RNAB gene of Saccharomyces cerevisiae, detection of
            the RNAB protein, and demonstration that it is essential for
            nuclear pre-mRNA splicing
            Mol. Cell. Biol. 8 (3), 1067-1075 (1988)
JOURNAL      88216580
MEDLINE
REFERENCE
AUTHORS      3 (bases 1 to 8151)
TITLE      Lossky,M., Anderson,G.J., Jackson,S.P. and Beggs,J.
            Identification of a yeast snRNP protein and detection of
            snRNP-snRNP interactions
            Cell 51 (6), 1019-1026 (1987)
JOURNAL      88080460
MEDLINE
REFERENCE
AUTHORS      4 (bases 1 to 8151)
TITLE      Hodges,P.E.
            Direct Submission
            Submitted (21-JUL-1993) Hodges P. E., University of Edinburgh,
            Institute of Cell and Molecular Biology, Darwin Building, King's
            Buildings, Mayfield Road, Edinburgh, Midlothian, United Kingdom,
            EH9 3JR
FEATURES
    source
        1. 8151
            /organism="Saccharomyces cerevisiae"
            /db_xref="taxon:4932"
            /chromosome="VIII, right arm"
            /clone="PR800"
            /clone_lib="Sau3A I genomic DNA in YEp24, D.Botstein
            (MIT)"
            1..1318
            /note="overlapping sequence with CDC23 gene (GenBank
            D00610)"
            431..7901
            /gene="PRP8"
            /note="5' and 3' ends mapped by nuclease S1 protection,
            contains no introns"
            /citation=[1]
            /citation=[2]
            /citation=[3]
            /db_xref="SGD:S0001208"
            /function="splicing factor, U5 snRNP protein"
            /evidence="experimental"
            431..7901
            /gene="PRP8"
            /db_xref="SGD:S0001208"
            531..7772
            /function="pre-mRNA splicing factor, U5 snRNP protein"
            /note="presumed C.elegans homologue encoded by C50C3.6
            (Genbank U14433)"
            /citation=[1]
            /citation=[2]
            /citation=[3]
CDS
gene

```

```

/codon_start=1
/db_xref="SGD:S0001208"
/evidence-experimental
/product="PRP8"
/protein_id="CAA80854.1"
/db_xref="pid:g395157"
/db_xref="GI:395157"
/db_xref="SWISS-PROT:P33334"
/translation="MSGLPPEPPGFEDSDLAPPPPPPPGCEIEIEIDNPVPSVN
EDTLPPEPPPSNFINAEIIVDTLPPEPPPGIDLETAKKVELHGRKLDIG
KDTETRRSKRRAKMTKRAKSNLYTPRAEPPHRLRIITVHSDMASKMNTRKA
FLGALKYPAHAIKILNMPHPMEQAKKVKLYHNSGATTFNPNPVPYEPETQWS
ATWLMREKBDRTKFRKMRPPPPDDPPSLSEPHNTEIPEDPINPLSDQDETY
KDWLDSRPLEEDSKVNGSTKMSFDPEKSNLYRUSTPLADEVTQNTYLYDKK
SFEKALNNAIPGPKFEPILPREEDYNGFNSIDRVIFRPVIRSEKVAPELHN
SRPSVRIPVNNPVCIIIONDEYDTPALFDPSPINPFIIDNNSLANTKENG
DTPLEDEAPLLAEELILPNTKAMSLYHSPFPNPKGMAVADYALAKWELQ
HPDEYPAVKVYQKLLKNVNLHPTLPNNHKLKSLKTKKYPQOTTIMVE
AGLOCRGCHNMLNLIHRKGLTYHLDYNNLPKPYKTLTTERKRSKLSGFHMR
LKKMKLVDTHVQFRLGNDAFQJLADGHTYLNHIGOLTGYRKYVMQIRACKD
LKHIYKFNKNLGPGCGEWPQPMRWMLNPLRGYIPLEXYIGNLITROGENSE
IVKTTKORLDAVYDLELNSYMDLILEMPEISIRQKARTILOHLSAMRCWKANIP
MDVPMAPIKKIIERYIKSKADAWSAHYRERIKGAHVEKTMVKNIGRLRLM
IKNEOROIQKNGPEITPEATTIFGSVMYMLRSRSPPIPPPLTYKNTKITVL
ALFDKDYVASKVRNASSREBALIIEAYNDPHDTLNRIKYLITQRYKRYDTTM
ENYONISPVISYDPLEKITDAYLDQYLNTEADQRLFPWIKRPLTQRYKRYDTTM
GINNLSEIMVSRGSAVLETTIGEMAEKIDFTLNRLIYDNPDIATKANVY
VINFEKMSHVNYGLIRGLKASFILOYGLVIDLLGQERATDLAPANNPEMO
FKSEVERAHPILRYTRYLRYMLFHEDEGEELTDEYLAENDPPNENIGNR
KMPKDSMRRLIRDVNIGRAVMEIOSRVPSTLSIKWENAFVYKKNPNLPSM
CGFEYRILIPROMEYVSNDEGVNDYDERTKQRAKAYIKISEIKFDSRTGIL
MASGTTTAAKNNLSISLFTYFREIATVEPLDILVAGEIRIQNRYVGLNSK
MPTRPAPVATPKELGGLGMSASHILIPASDLSMSKOTDGIHFRAGMHEDEKL
IPITERYITWENELDSQVAAEYATKROEAIQONRRLAFEELEGSDRGIPRISTL
FORDRHTLAYDRGHRIIRREFKOYSLERNPSHSDGKLMNNAVETDIOALG
IETILEHTLFGTGFNSGEMLEKASGPEDSMOKRGLSOIPNRPRTLM
SPTIRNAVYGVFIYOLDLGTGFLGKIPITLISIQIFRAHLMOKIESIYFDQOI
LDGELDVQIISVTKETVHPKSTIKNSASADITMSVHEMVSFSLHETNDFKG
LITNKMFDVOLRYGDYSDHSIRYRAKFLDYTDYDMSYSPSGVAGIDLAANY
DAYGEMNGKLPLONSMRTIKKAMPALVYLERIRKGIQSSVOEPELSSYAE
LENNDIKLFVDITNVYRVYHKTPEGNAKINCIETLNKTHLFLKITHSVMA
GOKRLSOLAKMTAEVSALVSLPKESOPQOIIYTRAMLODLEVHMLDEPNIAIRP
TELRPFSAASIDKLDVYKATKATPOVATRYTYDMDRISYTAFSRTITLALK
TNBSAKMILSDPTITTKSTHMLPSPFDEQITLESQRIILITEYCKINVISAL
TQTEIKDILIQNIKAPSVKQKMAELAAERKONDEAAAGATVMKTKINMAGEE
IVVVASADYESQTESKMNEMKSAIANTLLYRLNIVYASADFEQONVYLPNLL
KKEISDVKIOVAFIYGMSAKDHPKVEIKETVYVQLGVGSVOISNIDIDLP
DTEGELGEMHTOREELKPMASAVATSHSKFADKRDCCIISIFSGVSLAVN
LTDGEOYKGEENKIDIMNLSBGEPTSTHALLSDRTGCFITISFGVNYTTMGT
APNODGDIYFKIPLFTYNEHMRPVHFLQSELAGDELEAEOIDVFS"
7570
/gene="prp8-1"
7570
/gene="prp8-1"
7570
/note="type SRP8.3"
/citation=[1]
/citation=[2]
/replace="a"
/phenotype="temperature-sensitive pre-mRNA splicing
defect"
7856..7859
/gene="PRP8"
/note="major site, mapped by nuclease S1 protection"
/citation=[1]
/db_xref="SGD:S0001208"
/evidence-experimental
7863
/gene="PRP8"
/note="minor site, mapped by nuclease S1 protection"
/citation=[1]
/db_xref="SGD:S0001208"
/evidence-experimental
7866
3'polyA_site

```

```

Query Match
Best Local Similarity 94.7% Pred No.2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0;
Gaps 0;

QY 2 TATTTCATGAACAAATGA 20
Db 5677 TATTACATGAACAAATGA 5695
|||||
RESULT 9
YSCDBF3A YSCDBF3A 7845 bp DNA PLN 22-MAY-1995
LOCUS YSCDBF3A 7845 bp DNA PLN 22-MAY-1995
DEFINITION YSCDBF3A 7845 bp DNA PLN 22-MAY-1995
ACCESSION L29421
NID 9460041
VERSION L29421.1 GI:460041
SOURCE Saccharomyces cerevisiae (library: genomic) DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE
1 (bases 1 to 7845)
Dol,A. and Dol,X.
Cloning and nucleotide sequence of the CDC23 gene of Saccharomyces
cerevisiae
JOURNAL Gene 91 (1), 123-126 (1990)
MEDLINE 90382687
REFERENCE
2 (bases 1 to 7845)
Shee,J.E., Toyn,J.H. and Johnston,L.H.
The budding yeast US snRP Prp8 is a highly conserved protein which
links RNA splicing with cell cycle progression
JOURNAL Nucleic Acids Res. 22 (25), 5555-5564 (1994)
MEDLINE 95140615
FEATURES
source
1..7845
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/tissue_lib="genomic"
/map="chromosome VIII right arm"
225..7466
/gene="DBF3"
/db_xref="SGD:S0001208"
225..7466
/gene="DBF3"
/note="ORF"
7866
/codon_start=1

```

```

/function="involved in the control of S phase and pre-mRNA
splicing"
/protein_id="Aa67044.1"
/db_xref="PID:9460042"
/db_xref="GI:460042"
/translation="MSGLPPLPPGFEEDSDLPAPPPPPGPEIEHLDPMWSSYN
EDFLPPPPPPPEFEINAEIYDFTLPPPPPPGDELDTKAKREKELGKKLDIG
KDTFVTRKSRKAKMTKAKRSNLTPKAPKPEHLKRTIINHSDASMYATDMS
FLGALKYLPKAIKLELMPHPEOAKEVVLHTSCALITFVEITPPVPTAOMS
ATWIAMREKRDTHFKRMFPEPDSEQLHSENEPELPIIDLPKDDDEYV
KMDYDRPLEEDSKKYNSTYKMSFDELSQHLSTPLDEVDLKKYLLFPK
SENGKALNNAITGCPKEEPLPREEEDNERNSTDRFVFRPSEYVAALITTI
PGVOCARHGIITICRYLSTMTSTRLAFPPSPAPRPIPHFDNNSLVASTKENG
DTLPEDFAPLAEELILPNTKDASLHSPPEPRKGNVRODVALKMWLQ
HPDEYVAVKVSYOKLKNVNLNEHPTPTNKNKLLKSLKNTKYFOQTIIDWE
AGLQCGGNMLNLIRKGLTYLHIDVFNKPTTKERKSLNSPHLME
LLKMKLIVDTHOFRGNDVADOLAGIYIINHIGQLGIRYKKNVHOIRACD
LKHIYKFNKNLKGCGCGFQPAWVWNLFRGTPILERYIGNITPREGSNE
IVTTKORLDAYDELRNSYVDILEMPEISROKATTIIOHSEAMCRKANIP
WDVGPAPITKTIERTKADAWSAAHYNERIKGAHVKTWKNLGLTILM
IKNEORQIOKNGEPITPEEATITFVAVELSESSEPIPEPLTYKNDKILVL
ALBDLDVYASKYRLNASEREELIEADNPHDILNRKYLLOVRYKPYDTIM
ENYONISPYVSPLEKTDAYLDQYLMYADOKLEPNRKSLSDSIPPLVYKMO
GINNSIEMCVSGOSAVILETLGEMAEKIDFTLNLRLITVDPNIATYKANNV
VINPKMSHVKNKGLRGKLPASPIROYGLVLDLILGERSDLADGAPNNEMEQ
FKSEYKAPIRILTRYLDRITMLTFHEDEBELDELTAENPDNPNFNSGYNNR
KCPKRSRMLIRQDVNLGRAVWEIOSRPTISLTKWNAVSVYSKNNPLLSM
CGEVRILPRORMEEVNSDEGWDLDETRKORTAKAYLVEEIEKEDSRIRGL
MAGSTFTFVAKWNLSLISLTFREAIATPEPLDILVKESTRIONKVLGNSK
MTRPPPAVYTPREKLGIMISASHILIPASLSNKGOTDITFRAGMHEDEKL
IPITFRITWMEFLDSORVMAEYATROEAIOONRRLAFEELESGMDGIPRISTL
FORDRHLATDRGHRIRERKQYSLERNSPFWNTNSHDKLNNAIRTDVIOALGG
LETLLEHTLEKGTGFNSWEGLFCEKASGEDSMQFKLTHAQRTGLSIPNRRFTLAW
SPTINRANVYVGLVOLDLGIFLHGKIPKLSLQIFPAHLMQKIHSEIVDIOI
LDGELVLOIESVTEKTHPRKSYKMSAADITMESVHEMESKSLHETDNRSG
LITNKMEFDVOLRYGDDSHDISRYVAKFLDYTDVNSVPTGTMIGIDLAYAM
DAYNGNENGLKPLIONSMPRTIMKAPALYRIRKGLQIYSSOPEPLNSNAY
LFWNDIKLPFDITNRYRYHAKTEEGVATKAINCGITFTNPTGHLTKITTSWA
GOKRLOLAWKTAEEVSALVRSLEEQKQIIVTRKALDELVAHMLDFPIALRP
TELRLPFSAMSIDKLDVWKATEPQWLFNTDMLDRISYAFSRLLTLRLAK
TNEESAKMILSPITTIKSYHLMPSTDEQMITESOMDLITTEGRYNNISAL
TOEIKDIIIGONIKAPSVKOKMAELEARSEKONDEBAAGASTWAKTKTINAGEE
IYVVASADVESOTFSSKNEMWRSKAIANTLLYTLKNIYVADPFVEONVYTRKML
KRTIESDVRIQVARTYKMSAKDRHRYKIKYVLPQGHGISOISITRIGCLP
DITGELGLWHTOTELKMASEVATVHKTLADKKRDCIDISISTPSVLSAVN
LTDGQWMEENKDDIMNVLSGEFPTSTHQAOLSDRIGNFIIPSGNWTYTFGT
AFNOEGDYNKYGIPLEFYEMHRPVHFLQFSELADDELEAQIDVFS"
BASE COUNT      2661 a 1456 c 1533 g 2195 t
ORIGIN

```

```

Query Match      87.0% Score 17.4; DB 7; Length 7845;
Best Local Similarity 94.7% Pred. No. 2.5e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      2 TATTTCATGAACAATGA 20
      11111111111111111111
Db      5371 TATTACATGAACAATGA 5389

RESULT 10
LOCUS   YSC9986      41664 bp      DNA      PLN      29-DEC-1997
DEFINITION Saccharomyces cerevisiae chromosome VIII cosmid 9986.
ACCESSION U00027 U00093
NID      9551319
VERSION  U00027.1 GI:551319
KEYWORDS
SOURCE   Baker's yeast strain-S288C (AB972).
          Saccharomyces cerevisiae
          Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
          Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 41664).

```

AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	REFERENCE								
Johnston, M., Andrews, S., Brinkman, R., Cooper, J., Ding, H., Dover, J., Du, Z., Favell, A., Fulton, L., Gattung, S., Geisel, C., Kirsten, J., Kucaba, T., Hillier, L., Jier, M., Johnston, L., Keppeler, D., Langston, Y., Latreille, P., Louis, E., Macri, C., Mardis, E., Mousier, L., Nhan, M., Rifken, L., Riles, L., St. Peter, H., Thornton, L., Trevisakis, E., Vaubin, M., Vaughan, K., Vignati, D., Wilcox, L., Willis, A., Wilson, R., Wohlman, P., and Waterston, R.	Complete nucleotide sequence of Saccharomyces cerevisiae chromosome VIII	Science	265 (5181), 2077-2082 (1994)	94378003	2 (bases 1 to 41664)	Macri, C.	The sequence of S. cerevisiae cosmid 9986	Unpublished (1994)	3 (bases 1 to 41664)	Waterston, R.	Submitted	Submitted (26-FEB-1994)	4 (bases 1 to 41664)	Jia, Y. and Cherry, J.M.	Submitted	Submitted (29-AUG-1997)	Department of Genetics, Stanford University, Saccharomyces Genome Database, Stanford, CA 94305-5120, USA	On Sep 29, 1994 this sequence version replaced gi:458890.	Curated by: Saccharomyces Genome Database URL: http://genome-www.stanford.edu/ e-mail: yeast-curator@genome.sta.nford.edu								
Submitted by: Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA e-mail: mjessequence@wustl.edu	NEIGHBORING COSMID INFORMATION:	YSC9986 overlaps cosmid submissions YSCHL4005 and YSCH9186 by 200 bases.	Location/Qualifiers	1..41664	/organism="Saccharomyces cerevisiae"	/db_xref="taxon:4932"	/map="VIII"	/strain="S288C (AB972)"	862..1251	/gene="YHR162w"	862..1251	/note="Some similarity to Rat brain 0-44 mRNA, segment 2 (gp359720)"	/codon_start=1	/evidence-not_experimental	/product="Yhr162w"	/protein_id="AAB68009.1"	/db_xref="PID:9458905"	/translation="MTSSVRAFRFRFQSETPYVFWAPTLKMGVLVPAFGSDMKR PVEKISGAONLSLSTALWTRWFSVIRPNLLASVNSFCLTAGYOLGRYANRIR NDSISOLCSYILSGADESKKEITGR"	1420..2262	/gene="SOL3"	1420..2262	/gene="SOL3"	/note="YHR163W: Weak multicopy suppressor of los1-1"	/codon_start=1	/evidence-not_experimental	/product="SOL3p"	/protein_id="AAB68008.1"

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 1044017. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true left end of clone 50A13 (Z92545) is at 122226 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at

<http://www.sanger.ac.uk/HGP/CHX>
1044017 is from the library RPlC15 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:PCYPAC2>.

FEATURES

```
source
1..122325
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="1044017"
/map="p11.3-11.4"
/clone.lib="RPlC15"
382..2592
/feature="LI repeat: matches 2060. .4298 of consensus"
repeat_region
2788..2815
/note="14 copies 2 mer ta 89% conserved"
repeat_region
3426..3539
/note="LI repeat: matches 5191. .4664 of consensus"
repeat_region
3555..3890
/note="LI repeat: matches 5214. .5330 of consensus"
repeat_region
4119..4351
/note="LI repeat: matches 381. .586 of consensus"
repeat_region
4430..8034
/note="LI repeat: matches 589. .4151 of consensus"
repeat_region
6030..10187
/note="LI repeat: matches 3232. .5390 of consensus"
repeat_region
10039..10924
/note="LI repeat: matches 1. .908 of consensus"
repeat_region
10931..11389
/note="LI repeat: matches 75. .529 of consensus"
repeat_region
11390..11692
/note="Alu repeat: matches 1. .301 of consensus"
repeat_region
11708..12073
/note="LI repeat: matches 516. .890 of consensus"
repeat_region
12081..14318
/note="LI repeat: matches 3060. .5390 of consensus"
repeat_region
14201..15013
/note="LI repeat: matches 30. .872 of consensus"
repeat_region
16051..16359
/note="MER3 repeat: matches 323. .1 of consensus"
repeat_region
17313..17398
/note="MIR repeat: matches 125. .212 of consensus"
repeat_region
17490..18370
/note="LI repeat: matches 897. .1 of consensus"
repeat_region
18221..20110
/note="LI repeat: matches 5390. .3513 of consensus"
repeat_region
20133..20205
/note="MIR repeat: matches 316. .388 of consensus"
repeat_region
20134..20205
/note="MIR repeat: matches 402. .453 of consensus"
repeat_region
20243..20391
/note="Alu repeat: matches 146. .1 of consensus"
```

```
repeat_region
20515..21058
/note="LI repeat: matches 904. .356 of consensus"
repeat_region
21058..21632
/note="LI repeat: matches 3898. .4474 of consensus"
repeat_region
24600..24811
/note="MIR repeat: matches 18. .233 of consensus"
repeat_region
24925..25073
/note="MIR repeat: matches 380. .529 of consensus"
repeat_region
25572..25733
/note="MIR repeat: matches 221. .41 of consensus"
repeat_region
28048
/note="MIR repeat: matches 1. .424 of consensus"
repeat_region
29542
/note="LI repeat: matches 757. .1079 of consensus"
repeat_region
30245..30463
/note="MIR repeat: matches 260. .35 of consensus"
repeat_region
31128
/note="MIR repeat: matches 12. .254 of consensus"
repeat_region
31940..32320
/note="MIR repeat: matches 1. .426 of consensus"
repeat_region
32748..33119
/note="MIR repeat: matches 1712. .2086 of consensus"
repeat_region
33216..33503
/note="LI repeat: matches 1695. .1988 of consensus"
repeat_region
33779..34487
/note="LI repeat: matches 204. .917 of consensus"
repeat_region
34896..34941
/note="23 copies 2 mer tg 96% conserved"
repeat_region
35711
/note="LI repeat: matches 867. .101 of consensus"
repeat_region
36011
/note="Alu repeat: matches 300. .1 of consensus"
repeat_region
39594
/note="LI repeat: matches 5356. .1821 of consensus"
repeat_region
40328
/note="LI repeat: matches 1294. .631 of consensus"
repeat_region
40710..41173
/note="match: GSS A0017146 clone 2302K13"
repeat_region
41281..41417
/note="LI repeat: matches 919. .1055 of consensus"
repeat_region
41427..42116
/note="LI repeat: matches 2155. .2869 of consensus"
repeat_region
42137..42434
/note="Alu repeat: matches 1. .299 of consensus"
repeat_region
42458..42891
/note="LI repeat: matches 2906. .3339 of consensus"
repeat_region
42909..43377
/note="LI repeat: matches 3479. .3945 of consensus"
repeat_region
43355..43777
/note="LI repeat: matches 675. .539 of consensus"
repeat_region
44532..44707
/note="LI repeat: matches 673. .922 of consensus"
repeat_region
44600..45088
/note="MIR repeat: matches 1. .186 of consensus"
repeat_region
45124..45227
/note="MIR repeat: matches 69. .541 of consensus"
repeat_region
45442..45778
/note="MIR repeat: matches 488. .377 of consensus"
repeat_region
45567
/note="MIR repeat: matches 184. .102 of consensus"
repeat_region
46102
/note="25 copies 2 mer at 80% conserved"
repeat_region
46718..47172
/note="match: GSS A0004493 clone 2294C7; match: GSS A0006878 clone 2294C10"
repeat_region
46787..46898
/note="MIR repeat: matches 152. .262 of consensus"
repeat_region
47199..47458
/note="LI repeat: matches 642. .904 of consensus"
repeat_region
48589..48663
/note="MIR repeat: matches 120. .46 of consensus"
```



```
repeat_region complement(49037..49199)
repeat_region /note="L1M1 repeat: matches 661..495 of consensus"
repeat_region 49033..50335
repeat_region /note="L1M1C repeat: matches 28..466 of consensus"
repeat_region complement(50441..50744)
repeat_region /note="Alu10 repeat: matches 238..1 of consensus"
repeat_region 50803..50826
repeat_region /note="12 copies 2 mer ta 96% conserved"
repeat_region 50844..51059
repeat_region /note="L1M4A repeat: matches 834..1047 of consensus"
repeat_region 51090..51719
repeat_region /note="L1 repeat: matches 4760..5390 of consensus"
repeat_region 51575..52464
repeat_region /note="L1P15 repeat: matches 1..897 of consensus"
repeat_region 52842..53043
repeat_region /note="L1P13 repeat: matches 697..899 of consensus"
repeat_region 53910..53978
repeat_region /note="L1M2 repeat: matches 971..1044 of consensus"
repeat_region complement(53981..54055)
repeat_region /note="MSTR repeat: matches 424..346 of consensus"
repeat_region complement(54054..54363)
repeat_region /note="MSTR repeat: matches 310..1 of consensus"
repeat_region 54889..55140
repeat_region /note="MIR repeat: matches 2..261 of consensus"
repeat_region complement(55862..55914)
repeat_region /note="L1M1 repeat: matches 517..464 of consensus"
repeat_region 57446..57479
repeat_region /note="17 copies 2 mer ca 88% conserved"
repeat_region 57483..57522
repeat_region /note="20 copies 2 mer ga 90% conserved"
repeat_region 58284..58487
repeat_region /note="L1 repeat: matches 1761..1981 of consensus"
repeat_region 58677..59040
repeat_region /note="THEIC repeat: matches 3..371 of consensus"
misc_feature <58820..59202
/note="match: STS G13783"
```

```
Query Match 87.0%; Score 17.4; DB 9; Length 122325;
Best Local Similarity 94.7%; Pred. No. 2.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
OY 2 TATTTCATGAACAATGA 20
||||| ||||||| |||||||
Db 122021 TATTTCATGAACAATGA 122039
```

```
RESULT 12
LOCUS HS37M17 101076 bp DNA PRI 08-AUG-1996
DEFINITION Human DNA sequence from PAC 37M17 chromosome X.
ACCESSION #278022
NID 91486355
VERSION 278022.1 GI:1486355
KEYWORDS X.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
```

```
REFERENCE 1 (bases 1 to 101076)
```

AUTHORS

JOURNAL

COMMENT Direct Submission
Submitted (07-AUG-1996) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, UK. E-mail enquiries: humquerry@sanger.ac.uk
IMPORTANT: This sequence is the entire insert of clone 37M17. The
right end of clone 37M17 is at 101076.
37M17 is from the human PAC library.

FEATURES

```
source 1..101076
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
```

```
/map="X"
/clone="37M17"
3..281
/partial
/note="Alu repeat: matches 289..1 of consensus"
2344..2515
/note="MER20 element fragment"
9407..9448
/note="21 copies of 2 mer 83 % conserved"
11371..11410
/note="10 copies of 4 mer 85 % conserved"
11998..12189
/note="MIR element fragment"
12476..12775
/partial
/note="Alu repeat: matches 308..1 of consensus"
12957..13031
/note="L1 element fragment"
14856..14893
/note="MERSA element fragment"
15443..15796
/note="2 copies of 177 mer 98 % conserved"
15860..16159
/partial
/note="Alu repeat: matches 308..1 of consensus"
16169..16267
/note="MIR element fragment"
17729..18030
/partial
/note="Alu repeat: matches 308..1 of consensus"
18690..19019
/note="MSTR element fragment"
19564..19919
/note="MER18 element fragment"
19845..19919
/note="MER1A element fragment"
20108..20175
/note="17 copies of 4 mer 91 % conserved"
20205..20485
/partial
/note="Alu repeat: matches 297..1 of consensus"
21383..21449
/note="L1 element fragment"
21494..21696
/note="L1 element fragment"
21751..22083
/note="L1 element fragment"
22141..22282
/partial
/note="Alu repeat: matches 308..165 of consensus"
22283..22356
/note="L1 element fragment"
22356..22496
/partial
/note="Alu repeat: matches 162..1 of consensus"
22515..22889
/note="L1 element fragment"
22892..23000
/note="MSTR element fragment"
22907..23005
/note="MSTR element fragment"
22918..23211
/note="THEIB element fragment"
23079..23258
/note="MSTR element fragment"
23384..23602
/note="L1 element fragment"
23772..23957
/note="L1 element fragment"
24136..24249
/note="L1 element fragment"
24312..25217
/note="L1 element fragment"
```

```

repeat_region 26752..27026
/note="MER42A element fragment"
repeat_region 26776..26875
/note="L1 element fragment"
repeat_region 27831..28139
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 29106..29376
/partial
/note="Alu repeat: matches 287..1 of consensus"
repeat_region 29400..29687
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 29927..30033
/note="MLT1A element fragment"
repeat_region 29943..30005
/note="MSTC element fragment"
repeat_region 30431..30722
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 32718..32789
/note="MST1A element fragment"
repeat_region 32733..32774
/note="TRHB1B element fragment"
repeat_region 32846..32896
/note="TRHB1B element fragment"
repeat_region 32923..33068
/note="MST1A element fragment"
repeat_region 32961..33043
/note="TRHB1B element fragment"
repeat_region 35360..41487
/note="L1 element fragment"
repeat_region 41609..41727
/note="MLT1D element fragment"
repeat_region 42134..42220
/note="MLT1E element fragment"
repeat_region 42604..44783
/note="L1 element fragment"
repeat_region 44819..45081
/partial
/note="Alu repeat: matches 21..308 of consensus"
repeat_region 45595..45894
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 46842..47139
/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 48908..49199
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 50812..50883
/note="MLT1B element fragment"
repeat_region 51397..51711
/note="MLT1B element fragment"
repeat_region 51835..52808
/note="L1 element fragment"
repeat_region 52869..52993
/partial
/note="Alu repeat: matches 8..145 of consensus"
repeat_region 52994..53107
/partial
/note="Alu repeat: matches 198..308 of consensus"
repeat_region 53138..53218
/note="L1 element fragment"
repeat_region 53226..53363
/partial
/note="Alu repeat: matches 1..144 of consensus"
repeat_region 55035..55097
/note="MLT1C element fragment"
repeat_region 55176..55292
/note="MLT1C element fragment"
repeat_region 55194..55292
/note="MLT1B element fragment"
repeat_region 56882..56944
/note="MER1B element fragment"
repeat_region 56975..57263

```

```

/partial
/note="Alu repeat: matches 308..1 of consensus"
repeat_region 57291..57427
/note="MER1B element fragment"
repeat_region 57354..57428
/note="MER1A element fragment"
repeat_region 57504..57783
/note="MER1B element fragment"
repeat_region 59781..60056
/partial
/note="Alu repeat: matches 306..1 of consensus"
repeat_region 60682..60973
/note="Alu repeat: matches 1..308 of consensus"
repeat_region 63108..63158
/note="MLT2C2 element fragment"
repeat_region 63339..63377
/note="MLT2C2 element fragment"
repeat_region 64695..64823
/note="L1 element fragment"
repeat_region 65160..65612
/note="L1 element fragment"
repeat_region 65665..67096
/note="L1 element fragment"
repeat_region 67152..67235
/note="L1 element fragment"
repeat_region 67286..68295
/note="L1 element fragment"
repeat_region 68390..68571
/note="L1 element fragment"

```

Query Match 87.0%; Score 17.4; DB 9; Length 101076;
 Best Local Similarity 94.7%; Pred. No. 2.1e-02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTTCATGAAACAATGA 20
 Db 68713 TATTTCATGAAACAATGA 68695

```

RESULT 13
LOCUS AF111168/c DNA PRI 26-JAN-1999
DEFINITION Homo sapiens serine palmitoyl transferase, subunit II gene,
ACCESSION AF111168
VERSION 94186181
KEYWORDS AF111168.1 GI:4186181
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 231464)
AUTHORS Multimegabase Sequencing Group.
TITLE Complete sequence of the gene for serine palmitoyltransferase,
subunit II found on human chromosome 14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 231464)
AUTHORS UW Multimegabase Sequencing Group.
TITLE Direct Submission
JOURNAL Submitted (27-DEC-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, Washington 98195, USA
COMMENT On Jan 26, 1999 this sequence version replaced g1:4063505.
Sequencing methodology: high redundancy shotgun with plasmids.
Interspersed Repeats were identified with RepeatMasker (available
from http://ftp.genome.washington.edu/RM/RepeatMasker.html).

```

FEATURES

```

source
1..231464
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/map="14q24.3-31"
/clone="BAC 50114"

```

```
repeat_region complement(250.551) /clone.lib="RBC1.11"
repeat_region /rpt_family="AluY" complement(612.683)
repeat_region /rpt_family="(TAAAA)n" 722.1026
repeat_region /rpt_family="AluSg" 1027.1050
repeat_region /rpt_family="AT-rich" complement(1055.1496)
repeat_region /rpt_family="L1" complement(1680.1759)
repeat_region /rpt_family="MADE1" complement(2044.2343)
repeat_region /rpt_family="AluSg" 2345.2520
repeat_region /rpt_family="L1ME2" 2521.2819
repeat_region /rpt_family="AluY" 2820.3160
repeat_region /rpt_family="L1ME2" 3833.4128
repeat_region /rpt_family="AluSx" complement(4777.5064)
repeat_region /rpt_family="AluSx" 6726.7021
repeat_region /rpt_family="AluSx" complement(7166.7287)
repeat_region /rpt_family="FLAM_A" 7664.7966
repeat_region /rpt_family="AluSg" 7967.7994
repeat_region /rpt_family="(GAAA)n" complement(7997.8277)
repeat_region /rpt_family="MER57A" complement(8289.8569)
repeat_region /rpt_family="AluSg" complement(9087.9387)
repeat_region /rpt_family="AluSx" complement(9414.9521)
repeat_region /rpt_family="LINE2" complement(9547.9591)
repeat_region /rpt_family="(TGA)n" complement(9595.9814)
repeat_region /rpt_family="MIR" complement(10165.10287)
repeat_region /rpt_family="MIR" complement(10978.11185)
repeat_region /rpt_family="MIR" complement(11283.11475)
repeat_region /rpt_family="MIR" join(12291.12422,31490.31684,49761.49915,51955.52103,58362.58486,66381.66474,71724.71829,73351.73570,76648.76774,107289.107424,110703.110832,116467.116586)
note="intron-exon boundaries defined in relation to human cDNA sequence found in Y08686"
codon_start=1
product="serine palmitoyl transferase, subunit II"
protein_id="A00962.1"
db_xref="PID:94186182"
db_xref="GI:1186182"
translation="MRPEGGCCCRRTVRANGCVANGFVRNGYVRSSAAAAAAGG
IHHTONGGLKRPENAEETPMVLAVTVGVGLTFGLYLDPLKRYMRLEKHNH
TEREOKDFVLYODFENFTYRNLVYMRIRDMNRPICVPGARDIMEROHDYMSF
KYTGNIITKGVINMGSYNVLGFARNTGSGOEAAAYLVEYGAGVCSTREIGNLKEE
LEELVARTGVAAAMAYGMRATNSMNPALVAGKCLILSELNHSIVLARRSGAT
IRTRHNKMSLEKLDKDAIVYGPRTRRPMKKILILVEGITYSMGSLVRPEVIALK
KKRAYAYLDEAHSIGAAGTGRGVEYFGDPEVDVMTGTFKSGASGAYGGKK
ELIDYLRTHSHSAVATSLSPVVEOITTSKMGQDGTSLGKCVQOLAEINTYFR
RLKEMGFIYVGNDSPPVPLMLVPAKIGAFGRMLKRNIGVVVVGPPATPESRA
RFGLSAHTKEIILDTALKIDEVGLDLQKTSRHLNVLDRPDEITETED"
repeat_region 12430.12532
/rpt_family="(GGG)n"

repeat_region complement(13881.14162)
repeat_region /rpt_family="AluSg" complement(14185.14487)
repeat_region /rpt_family="AluY" complement(16941.16962)
repeat_region /rpt_family="AT-rich" complement(17032.17326)
repeat_region /rpt_family="AluSx" complement(17395.17556)
repeat_region /rpt_family="AluUo" complement(17557.17858)
repeat_region /rpt_family="AluY" complement(17863.17915)
repeat_region /rpt_family="FLAM_A" 17925.17970
repeat_region /rpt_family="(TAAA)n" complement(17991.18106)
repeat_region /rpt_family="FLAM_C" 18100.18250
repeat_region note="low quality data" complement(18108.18178)
repeat_region /rpt_family="FLAM_C" 18285.18415
repeat_region /rpt_family="FLAM_C" 18429.18729
repeat_region /rpt_family="AluSx" 18730.18979
repeat_region /rpt_family="L1MD2" complement(18991.19128)
repeat_region /rpt_family="AluUo" complement(19130.19415)
repeat_region /rpt_family="AluSx" 19419.19551
repeat_region /rpt_family="L1MB5" 19738.19908
repeat_region /rpt_family="LINE2" complement(19909.19933)
repeat_region /rpt_family="(TAA)n" complement(19934.20096)
repeat_region /rpt_family="AluSg" 20281.20367
repeat_region /rpt_family="LINE2" complement(21424.21529)
repeat_region /rpt_family="MADE1" 21459.21529
repeat_region /rpt_family="MADE1" complement(21869.22165)
repeat_region /rpt_family="AluUo" 22466.22496
repeat_region /rpt_family="(CA)n" complement(22524.22819)
repeat_region /rpt_family="AluSg" 23560.23858
repeat_region /rpt_family="AluUo" complement(23984.24284)
repeat_region /rpt_family="AluSg" 25864.25892
repeat_region /rpt_family="AT-rich" 25943.25996
repeat_region /rpt_family="L1MC4" 25997.26324
repeat_region /rpt_family="AluUo" complement(26420.26682)
repeat_region /rpt_family="LINE2" complement(26766.27294)
repeat_region /rpt_family="HERVH" complement(27297.27727)
repeat_region /rpt_family="LTR7" complement(28379.28340)
repeat_region /rpt_family="LINE2" complement(29470.29503)
repeat_region /rpt_family="AT-rich" complement(29935.30220)
```

Query Match	Best Local Similarity	87.0%	Score 17.4	DB 11	Length 23164
Matches 18	Conservative	0	Mismatches 1	Indels 0	Gaps 0
Db 21858	TATTTCATGAAACAATGA 20				
RESULT 14	LOCUS	CE56H9	28291 bp	DNA	INV 23-NOV-1998
DEFINITION	Caenorhabditis elegans cosmid F56H9, complete sequence.				
ACCESSION	274473				
NID	91405511				
VERSION	Z74473.1				
KEYWORDS	HTG; guanine nucleotide-binding protein.				
SOURCE	Caenorhabditis elegans.				
ORGANISM	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis				
REFERENCE	1 (bases 1 to 28291)				
AUTHORS	Burton, J.				
REFERENCE	Direct Submission				
TITLE	Submitted (20-JUN-1996) Louis, MO 63110, USA. E-mail: jee@sanger.ac.uk or tw@nemato.de.wustl.edu				
JOURNAL	2 (bases 1 to 28291)				
REFERENCE	Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copest, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardiner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wodlman, P.				
TITLE	2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans				
JOURNAL	Nature 368 (6466), 32-38 (1994)				
MEDLINE	94150718				
COMMENT	Coding sequences below are predicted from computer analysis, using predictions from GeneFINDER (P. Green, U. Washington), and other available information. For a graphical representation of this sequence and its analysis see: http://webpage.sanger.ac.uk/cgi-bin/play7db-wormceclass/sequence/subject-F56H9 Current sequence finishing criteria for the C. elegans genome				

[illegible]


```
misc_feature 16467..16583
              /note="exon predicted by xgrail, quality good"
              join(<16828..17139,17373..>18041)
mrna          /gene="T1B8.29"
              <16828..>18041
              /gene="T1B8.29"
gene          /note="predicted by genscan and genefinder"
              join(16828..17139,17373..18041)
CDS           join(16828..17139,17373..18041)

Query Match      85.0%; Score 17; DB 8; Length 114144;
Best Local Similarity 100.0%; Pred. No. 3.1e+02;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAAA 17
    |||||
Db 85586 GTATTTCATGAACAAA 85602
```

Search completed: September 13, 1999, 15:55:44
Job time: 4572 sec

THIS PAGE BLANK (USPTO)


```

FT      /note= "PI3- kinase p110"
PN      WO9321328-A.
PD      28-OCT-1993.
PF      13-APR-1993; G00761.
PR      13-APR-1992; GB-008135.
PA      (LUDM-) LUDWIG INST CANCER RES.
PI      Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI      ParkerPJ, Volinia S, Waterfield MD;
PI      WPI: 93-351738/44.
DR      P-PSDB; RA3341.
PT      Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT      activity, useful for controlling cell proliferation
PS      Claim 7; fig 9: 146pp. English
CC      Southern blot analysis was performed using a bovine cDNA probe contg.
CC      a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
CC      from a cDNA library constructed from mRNA isolated from the human
CC      cell line KG1a. Positive clones were sequenced to give the human
CC      PI3 kinase p110 sequence shown. This sequence has 95 percent
CC      homology with the bovine sequence. The domain encoding residues 19-
CC      100 of human p110 is sufficient to encode the kinase which will
CC      associate with the p85 kinase subunit. The gene may be used to
CC      provide a protein with PI3 kinase activity, and is useful for
CC      screening for (ant)agonists of PI3 kinase activity which could be
CC      useful for stimulation or inhibition of cell proliferation and hence
CC      prophylaxis or therapy. Platelet or neutrophil activity or blood
CC      glucose levels can be controlled using the kinase.
CC      See also 051156, 059012-23 and 057522-3.
SQ      Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match      100.0%; Score 20; DB 1; Length 3207;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GATTTCATGAACAATGA 20
DB      3111 GATTTCATGAACAATGA 3130

RESULT 3
O59021 059021 standard; DNA; 20 BP.
..      059021:
DT      12-APR-1994 (first entry)
DE      Sense PCR primer for p110.
KW      Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW      antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW      platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW      ss; amplification; p110.
CC      Synthetic.
CC      WO9321328-A.
PD      28-OCT-1993.
PR      13-APR-1992; G00761.
PR      13-APR-1992; GB-008135.
PA      (LUDM-) LUDWIG INST CANCER RES.
PI      Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI      ParkerPJ, Volinia S, Waterfield MD;
PI      WPI: 93-351738/44.
PT      Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT      activity, useful for controlling cell proliferation
PS      Example 1; Page 41; 146pp. English.
CC      An SGBAF-1 cell line was established by transfection of bovine
CC      adrenal cortex zona fasciculata cells with pSVneo. For p110
CC      poly (A)+ RNA was reversed transcribed with psVneo. For p110
CC      an antisense primer. For DNA polymerisation a sense primer was added
CC      and PCR performed. The prod. was p110.
CC      See also 051155-6, 059012-23 and 057522-3.
SQ      Sequence 20 BP; 9 A; 2 C; 3 G; 6 T;

Query Match      100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY      1 GATTTCATGAACAATGA 20
DB      1 GATTTCATGAACAATGA 20

RESULT 4
O57012 057012 standard; cDNA to mRNA; 3498 BP.
ID      057012:
AC      057012:
DT      31-AUG-1994 (first entry)
DE      Ptdins 3-kinase 110 kd catalytic subunit cDNA.
KW      110 kd catalytic subunit; phosphatidylinositol 3-kinase;
KW      transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW      Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
OS      Bos taurus.
FH      Key
FH      cds      Location/Qualifiers
FT      1..3207
FT      /*tag= a
FT      /product= p110

WO9403609-A.
PD      17-FEB-1994.
PF      05-AUG-1993; G01651.
PR      05-AUG-1992; GB-016654.
PA      (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI      Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI      WPI: 94-065697/08.
DR      P-PSDB; RA6294.
DR      Eukaryotic cells transformed with mammalian phospholipid or
PT      protein kinase DNA - useful in assays for compounds involved in
PT      cell growth regulation and for treating cancers
PS      Disclosure: Fig 1; 71pp; English.
CC      This sequence encodes the 110 kd catalytic subunit of the phosphatidylinositol (Ptdins) 3-kinase. This sequence was transformed into
CC      Schizosaccharomyces pombe cells under the regulatory control of the
CC      nmt promoter in an embodiment of the invention. In the presence of
CC      thiamine the promoter is inactive and the cells carrying the Ptdins
CC      catalytic subunit plasmid grow as the parental strain. In the absence
CC      of thiamine the nmt promoter functions and the Ptdins 3-kinase
CC      catalytic subunit is induced. Ptdins activity is substantially
CC      increased under these conditions. Cells containing constructs such as
CC      this, are useful in assays for detecting compounds involved in cell
CC      growth regulation. It is also used as the basis for detecting
CC      compounds for treating cancers and the formation of blood vessel
CC      plaques.
SQ      Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match      100.0%; Score 20; DB 1; Length 3498;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 GATTTCATGAACAATGA 20
DB      3111 GATTTCATGAACAATGA 3130

RESULT 5
T99449 799449 standard; DNA; 10670 BP.
ID      T99449:
AC      T99449:
DT      11-MAY-1998 (first entry)
DE      Staphylococcus carnosus nitrite reductase, nitrate reductase,
KW      nitrite reductase; NIRD; nitrate reductase; NARGHJ; Cy5G;
KW      nitrate reduction; nitrite reduction; pollutant; biotransformation;
KW      water treatment; vegetable; ds.
OS      Staphylococcus carnosus DSM 10563.
FH      Key
FH      Location/Qualifiers
FT      93..98
FT      /*tag= a
FT      116..122
FT      /*tag= b

```

FT	RBS	167. .171
FT	/*tag= c	
FT	CDS	181. .2586
FT	/*tag= d	
FT	/function= nitrb gene	
FT	CDS	2589. .2903
FT	/*tag= e	
FT	/function= nird gene	
FT	CDS	2894. .3847
FT	/*tag= f	
FT	/function= cysg gene	
FT	3853. .3878	
FT	/*tag= g	
FT	-35_signal	4037. .4042
FT	/*tag= h	
FT	-10_signal	4057. .4062
FT	/*tag= i	
FT	RBS	4128. .4134
FT	/*tag= j	
FT	CDS	4140. .7814
FT	/*tag= k	
FT	/function= narc gene	
FT	CDS	7804. .9381
FT	/*tag= l	
FT	/function= harc gene	
FT	CDS	9374. .9949
FT	/*tag= m	
FT	/function= narc gene	
FT	CDS	9942. .10625
FT	/*tag= n	
FT	/function= narc gene	
PD	EP-805205-A1.	
PD	05-NOV-1997.	
PF	02-MAY-1996: 201224.	
PR	02-MAY-1996: EP-201224.	
PA	(NEST) SOC PROD NESTLE SA.	
PI	Fast B, Galer W, Goetz F, Lindgren P, Neubauer H,	
PI	Pantel I;	
PI	WPI: 97-529059/49.	
DR	P-PSDB: W26733-39.	
PT	Recombinant S. carnosus nitrate and nitrite reductase proteins -	
PT	useful for reducing nitrate and nitrite in nitrate-polluted materials	
CS	Claim 3, Page 15-25, 66pp: English.	
CC	This DNA sequence from <i>Staphylococcus carnosus</i> includes genes that	
CC	code for the 2 subunits (see W26733 and W26734) of nitrite reductase	
CC	NitRD, the 3 subunits (see W26735, W26737 and W26738) of nitrate	
CC	reductase NARHJ, the nitrite reductase biogenesis protein NarI	
CC	(see W26739) and the nitrite reductase strom cofactor biosynthesis	
CC	protein CysG (see W26736). This gene region was identified	
CC	following 1991/ insertion mutagenesis studies. Claimed DNA	
CC	molecules (see T94449-51) can be used in the recombinant production	
CC	of claimed proteins (see W26733-49) involved in nitrate and nitrite	
CC	reduction in S. carnosus. These recombinant proteins, or cells	
CC	expressing them, in free or immobilised form, can be used in a	
CC	claimed method for reducing nitrate and nitrite in nitrate-polluted	
CC	water, vegetables and other polluted materials.	
SO	Sequence 10670 BP: 3350 A: 1771 C: 2304 G: 3045 T:	

```

Query Match      87.0%  Score 17.4;  DB 1;  Length 10670;
Best Local Similarity 94.7%  Pzed. No. 23;
Matches 18;  Conservative 0;  Mismatches 1;  Indels 0;  Gaps 0

QY      1  GTATTCATGAACAACATG 19
          ||| ||| ||| ||| ||| |||
DB      6284  GTATTCATGAACAACATG 6302

RESULT      6
Q27987
ID      Q27987  standard; DNA: 1586 BP.
AC      Q27987;
DT      -11-FEB-1993  (first entry) ,

```

QV	Sequence	1586 BP,	590 A;	302 C;	219 G;	475 T;
CC	Specifically cleaves the peptide bond at the C-terminus of the					
CC	glutamic acid residue in polypeptide					
CC						
CC	by PCR using the primer sequences given in Q21960-86. The protease					
CC	from <i>Staphylococcus aureus</i> strains. The DNA sequences were isolated					
CC	from <i>Staphylococcus aureus</i> strains.					
CC	The sequences given in Q21987-88 encode proteases which were isolated					
PS	Disclousure, Page 13-15; 25pp; Japanese.					
PT	acid residues in polypeptide(s)					
PT	capable of cleaving peptide bond at carboxyl terminus of glutamic					
PT	Novel protease prep. using <i>Bacillus</i> or <i>Saccharomyces</i> host -					
DR	P-PSDB; R26842.					
DR	WPI: 92-304938/37.					
FA	(SHIO) SHIONOGI & CO LTD.					
PR	20-FEB-1990; JP-040398.					
PD	03-AUG-1992.					
PN	J04211370-A.					
FT	mat_peptide	556.	1425			
FT		/*tag=	b			
FT	signal_peptide	352.	555			
FT		/*tag=	a			
FT		352.	1425			
OS	Staphylococcus aureus.					
OS	Protease; PCR; amplify; Staphylococcus; ss.					
KW	Protease from <i>S. aureus</i> ATCC12600.					
FT	key					
FT	Location/Qualifiers					

Query Match	84.0%	Score 16.8	DB 1	Length 1586
Best Local Similarity	90.0%	Pred. No. 40		
Matches 18	Conservative 0	Mismatches 2	Indels 0	Gaps 0
QY	1	GTATTTCATGAAACCAATGA	20	
DB	1072	GTATTTAATGAAAAAAATGA	1091	-

RESULT	7
ID	Q27988
AC	Q27988; standard; DNA; 1558 BP.
DI	11-FEB-1993 (first entry)
DE	Protease from S. aureus.
KW	Protease; PCR; amplify; staphylococcus; ss.
OS	Staphylococcus aureus.
FH	Key
FT	cds
FT	location/Qualifiers
FT	/tag= a
FT	signal_peptide 354..557
FT	/tag= b
FT	mat_peptide 558..1361
FT	/tag= c
PN	J04211370-A.
PD	03-AUG-1992.
PE	19-FEB-1991; 024633.
PR	20-FEB-1990; JP-040398.
PA	(SHRO) SHIONOGI & CO LTD.
DR	WPI: 92-304938/37.
PT	P-PDB: R29644.
CC	Novel protease prep'd. using Bacillus or Saccharomyces host - capable of cleaving peptide bond at carboxyl terminus of glutamic acid residues in polypeptide(s) Dislosure: Page 15-16; 25pp; Japanese. The sequences given in Q27987-88 encode proteases which were isolated from staphylococcus aureus strains. The DNA sequences were isolated by PCR using the primer sequences given in Q27960-86. The protease specifically cleaves the peptide bond at the C-terminus of the glutamic acid residue in polypeptide. Sequence 1558 BP; 579 A; 282 C; 221 G; 476 T;

```
Query Match      84.08; Score 16.8; DB 1; Length 1558;
```

Best Local Similarity 90.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GATTTCATGAACAATGA 20
||||| ||||| |||||

Db 1074 GATTTCATGAACAATGA 1093

RESULT 8

ID V75063 standard; DNA; 1019 BP.

AC V75063;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #752.

KM Computer readable medium; vaccine; S.aureus infection; immunodetection;

KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KM skin infection; surgical wound infection; scalded skin syndrome;

KM toxic shock syndrome; ds.

OS Staphylococcus aureus.

PH Key Location/Qualifiers

FT misc-feature 901..960

FT /*tag- a

FT /*note- "these bases represent a line of missing text in

the sequence listing in the specification. They

are included to maintain the nucleotide numbering

given in the specification for this DNA sequence

EP-786519-A2.

PD 30-JUL-1997.

PR 07-JAN-1997; 100117.

PR 05-JAN-1996; US-009861.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA; 97-374922/35.

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

PS Claim 1: Page 1646-1647; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

CC of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

CC the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or

CC industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can

CC be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,

CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock

CC syndrome, etc. Organisms transformed with the DNA sequences can be used

CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S.aureus DNA sequences contained on the

CC computer readable medium.

CC Sequence 1019 BP; 364 A; 194 C; 159 G; 240 T;

Query Match 84.0%; Score 16.8; DB 1; Length 1019;

Best Local Similarity 90.0%; Pred. No. 39;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GATTTCATGAACAATGA 20
||||| ||||| |||||

Db 666 GATTTCATGAACAATGA 685

RESULT 9

T23422/c standard; cDNA to mRNA; 255 BP.

AC T23422;

DT 16-AUG-1996 (first entry)

DE Human gene signature HUMGS05253.

KM Gene signature; messenger RNA; mRNA; relative abundance; frequency;

KM human; cloning; mapping; non-biased library; diagnosis; detection;

KM cell typing; abnormal cell function; ss.

OS Homo sapiens.

PN W09514772-A1.

PD 01-JUN-1995.

PE 11-NOV-1994; J01916.

PR 12-NOV-1993; JP-355504.

PA (MATS/) MATSUBARA K.

PA (OKUB/) OKUBO K.

PI Matsubara K, Okubo K;

DR WPI, 95-206931/27.

PT Identifying gene signatures in 3'-directed human cDNA library - e.g.

PT for diagnosis of abnormal cell function, by preparing cDNA that

PT reflects relative abundance of corresp. mRNA in specific human

PT tissues

PS Claim 1: Page 1365; 2245pp; Japanese.

CC A single-stranded DNA (or its complementary strand or the corresp.

CC double-stranded DNA) which comprises one of the 7837 "GS" sequences

CC given in T19001-T26837 and which is able to hybridise to part of

CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)

CC sequences were obtained from 3'-directed cDNA libraries prepared

CC from various human tissues; synthesis of cDNA was initiated from the

CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-

CC untranslated sequence is unique to a particular mRNA species, almost

CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library

CC is constructed so as to reflect accurately the relative abundance of

CC different mRNAs in the particular tissue from which it was derived.

CC The appearance frequency of a given GS in a cDNA library can be

CC determined (esp. using primers and probes derived from the GS

CC sequences) as a means of diagnosing abnormal cell function or for

CC recognising different cell types.

CC Sequence 255 BP; 84 A; 39 C; 31 G; 95 T;

Query Match 79.0%; Score 15.8; DB 1; Length 255;

Best Local Similarity 89.5%; Pred. No. 1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TATTTCATGAACAATGA 20
||||| ||||| |||||

Db 78 TATTTCATGAACAATGA 60

RESULT 10

T29074 standard; cDNA; 1796 BP.

ID T29074;

DT 13-SEP-1996 (first entry)

DE Arabidopsis adenylosuccinate lyase cDNA ADSL-1.

KM Adenylosuccinate lyase; ADSL; inhibitor; herbicide; ss.

OS Arabidopsis thaliana.

FS Key Location/Qualifiers

FT cds 14..1603

FT /*tag- a

FT misc-difference 1652

FT /*tag- b

FT /*note- "base 1652 is given as 'y' in the

FT specification"

PN W09618739-A1.

PD 20-JUN-1996.

PE 06-DEC-1995; E04796.

PR 14-DEC-1994; US-355770.

PR 26-OCT-1995; US-548509.

PA (CIBA) CIBA GEIGY AG.

PI Guyer CD, Ward ER;

DR WPI, 96-300656/30.

P-P-SDB; R97547.

PT DNA mol. encoding plant adenylosuccinate lyase (ADSL) - useful to

PT screen for inhibitors of ADSL, which may be used as herbicides

PS Claim 4: Page 13-17; 37pp; English.

CC A full-length cDNA (T29074), designated ADSL-1, codes for

CC adenylsuccinate lyase (ADSL) (R97547), an enzyme useful for
CC screening ADSL-inhibitors of potential use as herbicides. It was
CC isolated from an Arabidopsis thaliana lambda ZAP library using a cDNA
CC clone isolated by functional complementation of the E. coli purB
CC mutant JK268. A second cDNA fragment (T31801), ADSL-2, was also
CC isolated and represents a second distinct gene (773 sequence homology
CC to ADSL-1). ADSL-1 cDNA can be incorporated into a vector and used
CC for predn. of recombinant ADSL, or used as a probe to isolate
CC homologous genes, for gene mapping or for diagnosis of plant diseases
CC associated with ADSL.
SQ Sequence 1796 BP; 531 A; 336 C; 402 G; 526 T;

Query Match 79.0%; Score 15.8; DB 1; Length 1796;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GTATTCATGAACAATG 19
DB 502 GTATTCATGAACAATG 520

RESULT 11
V26289
ID V26289 standard; DNA: 3509 BP.
AC V26289;
DT 27-JUL-1998 (first entry)
DE Recombinant botulinum neurotoxin type B LH728/B encoding DNA.
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KM immunogen; detection; tetanus; non-toxic; toxin; ds.
OS Synthetic.
FH Clostridium botulinum.
FT Key Location/Qualifiers
FT CDS 1..3509
FT /*tag= a
FT /product= "LH728/B"
FT /note= "no stop codon given"

PN MO9807864-A1.
PD 26-FEB-1998.
PF 22-AUG-1997; G02273.
PR 13-DEC-1996; GB-025996.
PR 23-AUG-1996; GB-017671.
PA (MTCR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
PI Foster KA, Quinn CP, Shone CC;
PI WPI: 98-169168/15.
DR P-PSDB: W56017.
DR

PT Recombinant neurotoxin polypeptides - used to develop therapeutic
PT agents, immunogens or as non-toxic standards for the detection of
PT neurotoxins
PS Disclosure: Page 87-91; 137pp; English.
CC The present sequence encodes a recombinant neurotoxin protein from
CC +h3 present invention. The present invention describes recombinant
CC neurotoxin proteins which comprise a first and second domain, where
CC the first domain is adapted to cleave one or more vesicle or
CC plasma membrane associated proteins essential to exocytosis, and where
CC the second domain is adapted: (a) to translocate the protein into a
CC cell; (b) to increase the solubility of the protein compared to the
CC solubility of the first domain on its own, or (c) both to translocate
CC the protein into a cell and to increase the solubility of the protein
CC compared to the solubility of the first domain on its own, the protein
CC being free of clostridial neurotoxin (CN) and free of CN precursor that
CC can be converted into toxin by proteolytic action. The recombinant
CC proteins can be used as therapeutic agents for targeting cells
CC expressing a relevant substrate. The products can also be used as
CC immunogens and as non-toxic standards for the assessment and development
CC of in vitro assays for the detection of functional botulinum or tetanus
CC neurotoxins either in foodstuffs or in environmental samples.
SQ Sequence 3509 BP; 1468 A; 340 C; 548 G; 1153 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3509;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAACAATGA 20
DB 2344 TATTCATGAACAATGA 2362

RESULT 12
V26290
ID V26290 standard; DNA: 2574 BP.
AC V26290;
DT 27-JUL-1998 (first entry)
DE Recombinant botulinum neurotoxin type B LH417/B encoding DNA.
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KM immunogen; detection; tetanus; non-toxic; toxin; ds.
OS Synthetic.
FH Clostridium botulinum.
FT Key Location/Qualifiers
FT CDS 1..2574
FT /*tag= a
FT /product= "LH417/B"
FT /note= "no stop codon given"

PN MO9807864-A1.
PD 26-FEB-1998.
PF 22-AUG-1997; G02273.
PR 13-DEC-1996; GB-025996.
PR 23-AUG-1996; GB-017671.
PA (MTCR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.
PI Foster KA, Quinn CP, Shone CC;
PI WPI: 98-169168/15.
DR P-PSDB: W56018.
DR

PT Recombinant neurotoxin polypeptides - used to develop therapeutic
PT agents, immunogens or as non-toxic standards for the detection of
PT neurotoxins
PS Disclosure: Page 95-98; 137pp; English.
CC The present sequence encodes a recombinant neurotoxin protein from
CC +h3 present invention. The present invention describes recombinant
CC neurotoxin proteins which comprise a first and second domain, where
CC the first domain is adapted to cleave one or more vesicle or
CC plasma membrane associated proteins essential to exocytosis, and where
CC the second domain is adapted: (a) to translocate the protein into a
CC cell; (b) to increase the solubility of the protein compared to the
CC solubility of the first domain on its own, or (c) both to translocate
CC the protein into a cell and to increase the solubility of the protein
CC compared to the solubility of the first domain on its own, the protein
CC being free of clostridial neurotoxin (CN) and free of CN precursor that
CC can be converted into toxin by proteolytic action. The recombinant
CC proteins can be used as therapeutic agents for targeting cells
CC expressing a relevant substrate. The products can also be used as
CC immunogens and as non-toxic standards for the assessment and development
CC of in vitro assays for the detection of functional botulinum or tetanus
CC neurotoxins either in foodstuffs or in environmental samples.
SQ Sequence 2574 BP; 1060 A; 273 C; 403 G; 838 T;

Query Match 79.0%; Score 15.8; DB 1; Length 2574;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAACAATGA 20
DB 2344 TATTCATGAACAATGA 2362

RESULT 13
V26292
ID V26292 standard; DNA: 2574 BP.
AC V26292;
DT 27-JUL-1998 (first entry)
DE Recombinant botulinum neurotoxin type B LH417/B E.coli codon bias DNA.
KW Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KM immunogen; detection; tetanus; non-toxic; toxin; ds.

OS Synthetic.
 OS Clostridium botulinum.
 PN MO9807864-A1.
 PD 26-FEB-1998.
 PR 22-AUG-1997; G02273.
 PR 13-DEC-1996; GB-025996.
 PR 23-AUG-1996; GB-017671.
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PA (SPEY-) SPEYWOOD LAB LTD.
 PI Foster KA, Quinn CP, Shone CC;
 PI WPI: 98-169168/15.
 PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents, immunogens or as non-toxic standards for the detection of neurotoxins
 PT neurotoxins
 PS Example 2; Page 111-112; 137pp; English.
 CC The present sequence encodes LH417/B, a recombinant neurotoxin protein, with an Escherichia coli codon bias, from the present invention. The present invention describes recombinant neurotoxin proteins which comprise a first and second domain, where the first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and where the second domain is adapted: (a) to translocate the protein into a cell; (b) to increase the solubility of the protein compared to the solubility of the first domain on its own, or (c) both to translocate the protein into a cell and to increase the solubility of the protein compared to the solubility of the first domain on its own, the protein being free of clostridial neurotoxin (CN) and free of CN precursor that can be converted into toxin by proteolytic action. The recombinant proteins can be used as therapeutic agents for targeting cells expressing a relevant substrate. The products can also be used as immunogens and as non-toxic standards for the assessment and development of in vitro assays for the detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples.
 CC Sequence 2574 BP; 1038 A; 319 C; 412 G; 805 T;

Query Match 79.0%; Score 15.8; DB 1; Length 2574;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAAACAATGA 20
 ||||| ||||| |||||
 DB 2344 TATTTAATGAAAAAATGA 2362

RESULT 14
 V26293
 ID V26293 standard; DNA; 2574 BP.
 AC V26293:
 DT 27-JUL-1998 (first entry)
 DE Recombinant botulinum neurotoxin type B LH417/B encoding DNA.
 KM Botulinum; recombinant; Clostridium botulinum; neurotoxin;
 KW Immunogen; detection; tetanus; non-toxic; toxin; ds.
 OS Synthetic.
 OS Clostridium botulinum.
 PN MO9807864-A1.
 PD 26-FEB-1998.
 PR 22-AUG-1997; G02273.
 PR 13-DEC-1996; GB-025996.
 PR 23-AUG-1996; GB-017671.
 PA (MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PA (SPEY-) SPEYWOOD LAB LTD.
 PI Foster KA, Quinn CP, Shone CC;
 PI WPI: 98-169168/15.
 PT Recombinant neurotoxin polypeptides - used to develop therapeutic agents, immunogens or as non-toxic standards for the detection of neurotoxins
 PT neurotoxins
 PS Example 2; Page 112-114; 137pp; English.
 CC The present sequence encodes LH417/B, a recombinant neurotoxin protein, where bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B protein, from the present invention. The present invention describes recombinant neurotoxin

CC proteins which comprise a first and second domain, where the first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and where the second domain is adapted: (a) to translocate the protein into a cell; (b) to increase the solubility of the protein compared to the solubility of the first domain on its own, or (c) both to translocate the protein into a cell and to increase the solubility of the protein compared to the solubility of the first domain on its own, the protein being free of clostridial neurotoxin (CN) and free of CN precursor that can be converted into toxin by proteolytic action. The recombinant proteins can be used as therapeutic agents for targeting cells expressing a relevant substrate. The products can also be used as immunogens and as non-toxic standards for the assessment and development of in vitro assays for the detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples.
 CC Sequence 2574 BP; 1001 A; 466 C; 422 G; 685 T;

Query Match 79.0%; Score 15.8; DB 1; Length 2574;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAAACAATGA 20
 ||||| ||||| |||||
 DB 2344 TATTTAATGAAAAAATGA 2362

RESULT 15
 V30579
 ID V30579 standard; DNA; 3876 BP.
 AC V30579:
 DT 07-DEC-1998 (first entry)
 DE Clostridium botulinum type B toxin gene from Danish strain.
 KM Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
 KW botulism; BotB; ds.
 OS Clostridium botulinum serotype B Danish strain.
 PN MO9808540-A1.
 PD 05-MAR-1998.
 PR 28-AUG-1997; U15394.
 PR 28-AUG-1996; US-704159.
 PA (OPRI-) OPHIDIAN PHARM INC.
 PI Thallew BS, Williams JA;
 DR WPI: 98-230234/20.
 DR P-PSDB: W68392.
 PT Host cell containing recombinant expression vector encoding Clostridium botulinum type B or E toxin - useful to treat humans and other animals at risk of intoxication with clostridial toxin
 PS Example 35; Page 291-296; 428pp; English.
 CC This is the coding region of the Clostridium botulinum serotype B (Danish strain) toxin gene that codes for a 1291-amino acid polypeptide (see W68392). The C fragment (see W68394) of the B toxin has been expressed as histidine-tagged protein in Escherichia coli host cells. The invention relates to C. botulinum recombinant toxin polypeptides. Methods are provided which allow for the isolation of soluble recombinant proteins free of significant endotoxin contamination. Preferred hosts for production of the recombinant proteins are E. coli, insect cells and yeast cells.
 CC The recombinant proteins are used as immunogens for the production of vaccines and antitoxins that are useful in the treatment of humans and animals at risk of intoxication with clostridial toxin.
 CC Sequence 3876 BP; 1612 A; 370 C; 617 G; 1277 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3876;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TATTCATGAAACAATGA 20
 ||||| ||||| |||||
 DB 2344 TATTTAATGAAAAAATGA 2362

Thu Sep 16 12:23:39 1999

us-09-325-095-22.rng

Page 7

Search completed: September 13, 1999, 15:59:34
Job time: 4801 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds

(without alignments)
23.280 Million cell updates/sec

Title: US-09-325-095-22

Perfect score: 20
Sequence: 1 GATTTTCATGACAAATGA 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	20	100.0	292	23	D56232	D56232 HUM420G03B	
2	20	100.0	264	24	N23534	N23534 YV91G08..r1	
3	20	100.0	475	25	N44819	N44819 XY39A10..r1	
4	20	100.0	433	26	W76227	W76227 Z66A12..r1	
5	20	100.0	343	31	AA298585	AA298585 EST114216	
6	20	100.0	633	49	A1632172	A1632172 t885907.x	
7	19	95.0	241	25	N44835	N44835 YV39A10..r1	
8	17.4	87.0	367	34	AA499534	AA499534 v193h08.r	
9	17.4	87.0	285	49	A1643250	A1643250 v193h08.y	
10	16.8	84.0	472	30	AA242876	AA242876 z164d01.r	
11	16.8	84.0	360	36	C65080	C65080 C65080 f0j1	
12	16.8	84.0	398	42	A1099642	A1099642 34368 Lam	
13	16.4	82.0	264	21	F09317	F09317 HSC2Z112 n	
14	16.4	82.0	432	22	R44353	R44353 YV36B02..s1	
15	16.4	82.0	358	23	R87639	R87639 YP89C01..s1	
16	16.4	82.0	352	24	H82513	H82513 YS69B12..s1	
17	16.4	82.0	321	24	H94229	H94229 YV07F01..r1	
18	16.4	82.0	355	24	N21852	N21852 SMNHADA0180	
19	16.4	82.0	291	27	C05875	C05875 C05875 Huma	
20	16.4	82.0	730	34	C24336	C24336 C24336 Dict	
21	16.4	82.0	649	34	C24406	C24406 C24406 Dict	
22	16.4	82.0	343	36	AA632828	AA632828 np84d05.s	
23	16.4	82.0	481	38	AA796325	AA796325 vs94e08.r	
24	16.4	82.0	475	38	AA796972	AA796972 vs94e07.r	
25	16.4	82.0	293	43	A1168589	A1168589 ox11d04.s	
26	16.4	82.0	610	44	A0038541	A0038541 A0038541	
27	16.4	82.0	420	46	A1428066	A1428066 mm38b08.x	
28	16.4	82.0	1823	50	AF013092	AF013092 AF013092	
29	16	80.0	489	54	A1283067	A1283067 qm62c11.x	
30	16	80.0	556	51	A1727149	A1727149 BNLGH1745	
31	16	80.0	650	51	A1727527	A1727527 BNLGH1832	
32	16	80.0	743	51	A1729279	A1729279 BNLGH1130	
33	15.8	79.0	360	20	D36276	D36276 CELK030H7F	
34	15.8	79.0	294	20	T07560	T07560 EST05450 Fe	
35	15.8	79.0	324	20	T16380	T16380 NIB124 Nor	
36	15.8	79.0	248	20	T17087	T17087 NIB314 Norm	
37	15.8	79.0	331	20	T34362	T34362 EST6619 Hu	
38	15.8	79.0	459	20	T51830	T51830 YB54C02..s1	
39	15.8	79.0	275	20	Z41154	Z41154 HSC2P092 n	
40	15.8	79.0	448	22	H05472	H05472 Y181F06..s1	
41	15.8	79.0	413	22	H13474	H13474 Y115C06..r1	
42	15.8	79.0	331	22	R37868	R37868 YC98F08..s1	
43	15.8	79.0	231	22	R42218	R42218 YF90A11..s1	
44	15.8	79.0	363	22	R49580	R49580 YG68F11..s1	
45	15.8	79.0	375	51	A1716960	A1716960 UT-R-Y0-a	

ALIGNMENTS

RESULT 1
D56232/c D56232 292 bp mRNA EST 31-AUG-1995
LOCUS HUM420G03B Clontech human fetal brain polyA+ mRNA (#6535) Homo
DEFINITION sapiens cDNA clone GEN-420G03 5', mRNA sequence.
ACCESSION D56232
NID 9970756
VERSION D56232.1 GI:970756

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 292)
Fujitara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shiomiyu, H.,
Takahashi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
Maekawa, H., Shin, S. and Nakamura, Y.
Fujitara et al. (1995)
Unpublished (1995)
On Sep 21, 1992 this sequence version replaced gi:276155.

TITLE
JOURNAL
COMMENT

FEATURES
source
Contact: Tsutomu Fujitara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035
Insert Length: 770 Std Error: 0.00
High quality sequence stop: 331.
Location/Qualifiers
1..292
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1ib="Clontech human fetal brain polyA+ mRNA
(#6535)"

BASE COUNT
ORIGIN
78 a 59 c 55 g 100 t

Query Match
Best Local Similarity 100.0%; Score 20; DB 23; Length 292;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAAACAATGA 20
|||||

Db 116 GTATTTCATGAAACAATGA 97

RESULT 2
LOCUS N23534 264 bp mRNA EST 28-DEC-1995
DEFINITION yv91908.r1 Soares melanocyte 2NBHM Homo sapiens cDNA clone
IMAGE:250142.5' similar to SP:1100_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.
ACCESSION N23534
NID G1137684
VERSION N23534.1 GI:1137684
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 264)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The Washu-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:692929.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
High quality sequence stops: 225

Source: IMAGE Consortium, LNL.
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1135 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 225.
Location/Qualifiers
1..264
/organism="Homo sapiens"
/db_xref="GDB:3867848"
/db_xref="taxon:9606"
/map="14"
/clone_1ib="Soares melanocyte 2NBHM"
/clone_1ib="Soares melanocyte 2NBHM"
/issue_type="melanocyte"
/sex="Male"
/lab_host="PH10B (ampicillin resistant)"
/note="Vector: pT7/3D (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [3']
TGTACCAATCGAGTGGAGGAGCGCCAGTCTTTTCTTTTCTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT
ORIGIN
94 a 53 c 52 g 60 t

Query Match
Best Local Similarity 100.0%; Score 20; DB 24; Length 264;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAAACAATGA 20
|||||

Db 91 GTATTTCATGAAACAATGA 110

RESULT 3
LOCUS N44819 475 bp mRNA EST 13-FEB-1996
DEFINITION yy39a10.r1 Soares melanocyte 2NBHM Homo sapiens cDNA clone
IMAGE:273594.5' similar to SW:PILA_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
mRNA sequence.
ACCESSION N44819
NID G1185985
VERSION N44819.1 GI:1185985
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 475)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The Washu-Merck EST Project
Unpublished (1995)
On May 8, 1993 this sequence version replaced gi:800198.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.

Seq primer: T7
High quality sequence stop: 402.
Location/Qualifiers
1..475

FEATURES
Source
/organism="Homo sapiens"
/db_xref="GDB:3883236"
/db_xref="taxon:9606"
/clone="IMAGE:273594"
/clone_id="Soares melanocyte 2NDM"
/sex="male"
/tissue_type="melanocyte"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCCGACATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from normal foreskin melanocytes (FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 170 a 90 c 85 g 128 t 2 others
ORIGIN

Query Match 100.0%; Score 20; DB 25; Length 475;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAAATGA 20
|||||
DB 152 GTATTTCATGAACAAATGA 171

RESULT 4
LOCUS W76227 433 bp mRNA EST 16-OCT-1996
DEFINITION z664c12.r1 Soares_fetal_heart_NbH19W Homo sapiens cDNA clone IMAGE:345430 5' similar to SW:P11A_BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ; mRNA sequence.
ACCESSION W76227
VERSION g1386490
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 433)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,J., Hultman,M., Kucaba,T., Le,M., Lennon,G., Maira,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P., and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:693546.

TITLE JOURNAL
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1049 Std Error: 0.00
Seq primer: mob.REGA+ET
High quality sequence stop: 391.
Location/Qualifiers
1..433

/organism="Homo sapiens"
/db_xref="GDB:1270805"
/db_xref="taxon:9606"
/map="20"
/clone="IMAGE:345430"
/clone_id="Soares_fetal_heart_NbH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAGTGGAGCGCCGACATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbH19W."

BASE COUNT 148 a 78 c 90 g 116 t 1 others
ORIGIN

Query Match 100.0%; Score 20; DB 26; Length 433;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAAATGA 20
|||||
DB 276 GTATTTCATGAACAAATGA 295

RESULT 5
LOCUS AA298585 343 bp mRNA EST 18-APR-1997
DEFINITION EST114216 HSC172 cells II Homo sapiens cDNA 5' end similar to phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA sequence.
ACCESSION AA298585
VERSION g1950938
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 343)
AUTHORS Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Mai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodex,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Lin,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Sauder,D.M., Shirley,R., Small,K.Y., Spriggs,T.A., Uterback,T.R., Weidman,J.F., Li,Y., Bedarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dlake,D., Feng,D.-F., Ferrle,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,D., Li,H., Meisner,P.S., Olsen,H., Raymond,L., Wel,Y.F., Wing,J., Xu,C., Yu,G.-L., Ruben,S.M., Dillon,P.J., Pannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
On Sep 12, 1996 this sequence version replaced gi:1192803.
Other ESTs: THC168479
Contact: Kerlavage, AR

TITLE JOURNAL
MEDLINE
COMMENT
On Sep 12, 1996 this sequence version replaced gi:1192803.
Other ESTs: THC168479
Contact: Kerlavage, AR

Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavetlgr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/cdb/hg1/hg1.html>)
Seq primer: M13 Reverse.

FEATURES

source
1. 343
/organism="Homo sapiens"
/db_xref="taxon:9606"
/db_xref="taxon:9606"
/clone_lib="HSC172 cells II"
/cell_type="fibroblast"
/cell_line="HSC172 (60PDL)"
/dev_stage="fetal"
/note="Organ: lung; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT 132 a 61 c 64 g 84 t 2 others
ORIGIN

Query Match 100.0%; Score 20; DB 31; Length 343;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
|||||
Db 67 GTATTTCATGAACAATGA 86

RESULT 6
A1632172 633 bp mRNA EST 26-APR-1999
t585g07.x1 NCI_CGAP_GC6 Homo sapiens CDNA clone IMAGE:228108 3'
SIMILAR TO SW:P11A_BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE
CATALYTIC SUBUNIT, ALPHA ISOFORM ;, mRNA sequence.
ACCESSION A1632172
NID 94683502
VERSION A1632172.1 GI:4683502
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 633)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL On May 7, 1998 this sequence version replaced gi:3121411.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html

FEATURES

source
Seq primer: -40UP from Gibco
High quality sequence stop: 446.
Location/Qualifiers
1. 633
/organism="Homo sapiens"

/db_xref="taxon:9606"
/map="15q26.1"
/clone_image="228108"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="Pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pRT3D-Pac (Pharmacia) with a modified
polylinker. Plasmid DNA from the normalized library
NCI_CGAP_GC6 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (clones 1257096-1258631,
145064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 169 a 109 c 108 g 246 t
ORIGIN

Query Match 100.0%; Score 20; DB 49; Length 633;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GTATTTCATGAACAATGA 20
|||||
Db 440 GTATTTCATGAACAATGA 421

RESULT 7
M44835 241 bp mRNA EST 13-FEB-1996
Y939E10.r1 Soares melanocyte 2NDHM Homo sapiens CDNA clone
IMAGE:273642 5', similar to SW:P11A_BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
mRNA sequence.
ACCESSION M44835
NID 91186001
VERSION M44835.1 GI:1186001
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 241)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Roffling, T., Soares, M., Tan, F.,
Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
JOURNAL On May 8, 1995 this sequence version replaced gi:800214.
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: 77
High quality sequence stop: 1.
Location/Qualifiers
1. 241
/organism="Homo sapiens"
/db_xref="GDB:3883284"
/db_xref="taxon:9606"
/map="3 p24.2-p21.1"
/clone_image="IMAGE:273642"
/clone_lib="Soares melanocyte 2NDHM"

/sex="Male"
/tissue-type="melanocyte"
/lab-host="DH10B (ampicillin resistant)"
/note="vector: p773D (Pharmacia) with a modified
polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15'
TGTTCCATCATGATGAGGAGCGCCGACGTTTTTTTTTTTTTTT 3',
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified p773 vector
(Pharmacia). Library constructed by Bento Soares and
M.Fatima Bonaldo. RNA from normal foreskin melanocytes
(FS374) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 69 a 48 c 44 g 56 t 24 others

ORIGIN

Query Match 95.0%; Score 19; DB 25; Length 241;
Best Local Similarity 95.0%; Pred. No. 1.5e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GATTTTCATGAACAATGA 20
|||||
Db 148 GATATTCATGAACAATGA 167

RESULT 8
AA499534 367 bp mRNA EST 01-JUL-1997
LOCUS A499534
DEFINITION *v193h08.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:919839 5', mRNA sequence.
ACCESSION AA499534
NID 92234501
VERSION AA499534.1 GI:2234501
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 367)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1392797.

CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:532055
Seq primer: -28m3 rev1 ET from Amersham
High quality sequence stop: 345.
Location/Qualifiers
1..367
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone_image="919839"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:

ECORI: Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGTGT 3'."

BASE COUNT 100 a 86 c 98 g 83 t

ORIGIN

Query Match 87.0%; Score 17.4; DB 34; Length 367;
Best Local Similarity 94.7%; Pred. No. 5.4e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 TATTCATGAACAATGA 20
|||||
Db 154 TATTCATGAACAATGA 172

RESULT 9
A1643250 285 bp mRNA EST 29-APR-1999
LOCUS A1643250
DEFINITION v193h08.y1 Stratagene mouse heart (#937316) Mus musculus cDNA clone
IMAGE:919839 5', mRNA sequence.
ACCESSION A1643250
NID 94721725
VERSION A1643250.1 GI:4721725
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 285)
Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.

TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2980088.
CONTACT: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:532055
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 272.
Location/Qualifiers
1..285
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/map="7"
/clone_image="919839"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
ECORI: Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'."

FEATURES

source
1..367
/organism="Mus musculus"
/strain="NIH/Swiss"
/db_xref="taxon:10090"
/clone_image="919839"
/clone_lib="Stratagene mouse heart (#937316)"
/sex="pooled"
/tissue_type="heart"
/dev_stage="13 day embryos"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: heart; Vector: pBluescript SK-; Site_1:
ECORI: Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3'."

BASE COUNT 84 a 59 c 76 g 66 t
 Cytosin

Query Match 87.0%; Score 17.4; DB 49; Length 285;
 Best Local Similarity 94.7%; Pred. No. 6.1e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTTCATGAACAATGA 20
 ||| ||||| ||||| |||||
 Db 154 TATCTCATGAACAATGA 172

RESULT 10
 AA242876 472 bp mRNA EST 06-AUG-1997
 LOCUS zrf6401.r1 Soares.NhHMPu_S1 Homo sapiens cDNA clone IMAGE:668161
 DEFINITION 5', mRNA sequence.
 ACCESSION AA242876
 NID 91873669
 VERSION AA242876.1 GI:1873669
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Creniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 472)
 Hillier,L., Allen,M., Bowles,L., Dubnue,T., Gelsel,G., Jost,S.,
 Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J.,
 Moore,B., Schellenberg,K., Stepec,M., Tan,F., Theising,B.,
 White,Y., Wylie,T., Waterston,R. and Wilson,R.
 Washu-Merck EST Project 1997
 *Unpublished (1997)
 On May 5, 1995 this sequence version replaced gi:798270.

TITLE JOURNAL
 COMMENT
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert length: 761 Std Error: 0.00
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 460.
 Location/Qualifiers

FEATURES
 source
 1..472
 /organism="Homo sapiens"
 /db_xref="GDB:5562128"
 /db_xref="taxon:9606"
 /map="11"
 /clone="IMAGE:668161"
 /clone_lib="Soares.NhHMPu_S1"
 /tissue_type="Pooled human melanocyte, fetal heart, and
 pregnant uterus"
 /lab_host="DH10b"
 /note="Organ: mixed (see below); Vector: pT73D-Pac
 (Pharmacia) with a modified polylinker; Site: 1: Not I;
 Site: 2: Eco RI; Equal amounts of plasmid DNA from three
 normalized libraries (melanocyte 2NHM, pregnant uterus
 NHMPU, and fetal heart NBH11W) were mixed, and as circles
 were made in vitro. Following HAP purification, this DNA
 was used as tracer in a subtractive hybridization
 reaction. The driver was PCR-amplified cDNAs from pools of
 5,000 clones made from the same 3 libraries. The pools
 consisted of I.M.A.G.E. clones 260232-265223,
 340488-345479, and 484488-489479."

BASE COUNT 137 a 91 c 87 g 157 t
 ORIGIN

Query Match 84.0%; Score 16.8; DB 30; Length 472;

Best Local Similarity 90.0%; Pred. No. 8.4e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAATGA 20
 ||||| ||||| ||||| |||||
 Db 127 GTATTTCATGAACAATGA 146

RESULT 11
 C65080 360 bp mRNA EST 22-SEP-1997
 LOCUS C65080/c
 DEFINITION C65080 Yuii Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk39898 5', mRNA sequence.
 ACCESSION C65080
 NID 92423785
 VERSION C65080.1 GI:2423785
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabdilita;
 Rhabdilita; Rhabdilitidae; Rhabdilitidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 360)
 Kohara,Y., Mochizuki,T., Tabara,H., Watanabe,H., Sugimoto,A.,
 Sano,M., Miyata,A. and Nishigaki,A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 On Sep 12, 1996 this sequence version replaced gi:1402204.

TITLE JOURNAL
 COMMENT
 CONTACT: Yuii Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.
 Location/Qualifiers

FEATURES
 source
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 hlm-8(e1489)"
 /db_xref="taxon:6239"
 /clone="yk39898"
 /clone_lib="Yuii Kohara unpublished cDNA"
 /note="dev stage=varied, sex=Hermaphrodite male,
 tissue_type=whole animal."
 Location/Qualifiers

BASE COUNT 100 a 69 c 62 g 125 t 4 others
 ORIGIN
 Query Match 84.0%; Score 16.8; DB 36; Length 360;
 Best Local Similarity 90.0%; Pred. No. 9.5e+02;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTTCATGAACAATGA 20
 ||| ||||| ||||| |||||
 Db 183 GTCTTCTTCAACAATGA 164
 RESULT 12
 A1099642 398 bp mRNA EST 21-AUG-1998
 LOCUS A1099642/c
 DEFINITION A1099642 Lambda-PRL2 Arabidopsis thaliana cDNA clone 1607XP 3', mRNA
 sequence.
 ACCESSION A1099642
 NID 93449954
 VERSION A1099642.1 GI:3449954
 KEYWORDS EST.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsis.

REFERENCE 1 (bases 1 to 398)

AUTHORS Newman, T., deBruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomasnow, M., Retzel, E. and Somerville, C.

TITLE Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clones

JOURNAL MEDLINE Plant Physiol. 106, 1241-1255 (1994)

COMMENT On Jan 14, 1998 this sequence version replaced g1:1798073.

CONTACT: Thomas Newman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E. Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 2231jcn@lhm.cl.msu.edu

The sequence entry for this EST has been reverse complemented and is being submitted in the sense orientation.

Seq primer: M13-21.

Location/Qualifiers

FEATURES

source

1. 398

/organism="Arabidopsis thaliana"

/strain="var columbia"

/db_xref="taxon:3702"

/clone="160F7XP"

/note="Vector: lambda ZAP-Lox; Site_1: Sal; Site_2: Not; lambda PRL2 is a cDNA library derived from equal quantities of 4 pools of mRNA. The mRNA sources were 1) 7 day germinated etiolated seedlings; 2) tissue culture grown roots; 3) staged plants half with 24 hour light cycle, half on 16 hr light, 8 hour dark-roseettes; 4) same plants as 3 but aerial tissue (stems, flowers and siliques). The vector is BRU's lambda ZAP-Lox. The cDNA inserts were directionally cloned with Sal-Not arms using oligo dT primed cDNA."

BASE COUNT 100 a 102 c 70 g 123 t 3 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 42; Length 398;
Best Local Similarity 90.0%; Pred. No. 9e+02; 2; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTATTGATGAACAATGA 20
|||||

Db 305 GTATTGATGAACAATGA 286

RESULT 13

LOCUS F09317 264 bp mRNA EST 23-FEB-1995

DEFINITION HSC22H12 normalized infant brain cDNA Homo sapiens cDNA clone

ACCESSION F09317

NID 9678473

VERSION F09317.1 GI:678473

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 264)
Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houlgatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastien, K., Kachichis, C. and Tessier, A.

TITLE IMAGE: molecular integration of the analysis of the human genome and its expression

JOURNAL MEDLINE C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)

COMMENT 95277534

CONTACT: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, Bp60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

Single read. removed at sequence 5' end
Genexpress_Library_Idt: C; Genexpress_Sequence_Idt: alc-2zh11
Insert Length: 528 Std Error: 0.00
Seq primer: (-21)M13 universal
High quality sequence stop: 1.

Location/Qualifiers

FEATURES

source

1. 264

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="C-2zh11"

/clone.lib="normalized infant brain cDNA"

/sex="female"

/tissue_type="total brain"

/dev_stage="3 months old"

/note="Organ: brain; Vector: lambda BA; Site_1: HindIII; Site_2: NotI; sex: female; dev stage=3 months old; isolate=muscular atrophy patient; tissue_type=total brain; total mRNA was oligo(dT) primed and directionally cloned 5' -> 3' into the HindIII -> NotI sites of the lambda BA vector. Clone library from B.Scores, Psychiatry Dept. Columbia University, USA. Normalization method: Bento Scores, P.N.A.S. in press"

BASE COUNT 58 a 43 c 43 g 119 t 1 others

ORIGIN

Query Match 82.0%; Score 16.4; DB 21; Length 264;
Best Local Similarity 94.4%; Pred. No. 1.6e+03;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTTCATGAACAATGA 20
|||||

Db 103 ATTTCATGAACAATGA 86

RESULT 14

LOCUS R44353 432 bp mRNA EST 22-MAY-1995

DEFINITION YG36b02.s1 Scores infant brain INTB Homo sapiens cDNA clone

ACCESSION R44353

NID 9820649

VERSION R44353.1 GI:820649

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 432)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marie, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE The WashU-Merck EST Project

JOURNAL MEDLINE Unpublished (1995)

COMMENT On May 9, 1995 this sequence version replaced g1:803077.

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert Size: 1986

*High quality sequence stops: 292 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1986 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 292.
Location/Qualifiers

FEATURES

source

1.432
/organism="Homo sapiens"
/db_xref="GDB:406747"
/db_xref="taxon:9606"
/clone="IMAGE:34400"
/clone_1lb="Soares infant brain 1N1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lnfmd BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AAGTGAAGATTCGGCGCGGAGGATTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lnfmd BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 70 c 88 g 174 t 2 others

ORIGIN

Query Match

Best Local Similarity 94.4%; Score 16.4; DB 22; Length 432;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ATTTGATGAACAATGA 20
||||| |||||||

DB 113 ATTTGATGAACAATGA 96

RESULT 15

R87639 358 bp mRNA EST 16-AUG-1995
LOCUS YP89C01.s1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone
DEFINITION IMAGE:194592 3' similar to contains Alu repetitive element; mRNA
sequence.
R87639
R87639

ACCESSION R87639
NTD 9946452
VERSION R87639.1 GI:946452
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Atheria; Primates; Catarrhini; Hominoidea; Homo.
AUTHORS 1 (bases 1 to 358)
Haller, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hulman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Insert Size: 361
High quality sequence stops: 267
Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 361 Std Error: 0.00

FEATURES

source

Seq primer: Promega -21ml3
High quality sequence stop: 267.
Location/Qualifiers
1.358
/organism="Homo sapiens"
/db_xref="GDB:3763642"
/db_xref="taxon:9606"
/clone="IMAGE:194592"
/clone_1lb="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: p773D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AAGTGAAGATTAATTAAGATCTTTTCTTTTCTTTT 3'];
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p773 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 98 a 76 c 97 g 83 t 4 others

ORIGIN

Query Match

Best Local Similarity 94.4%; Score 16.4; DB 23; Length 358;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TATTTGATGAACAATG 19
||||| |||||||

DB 91 TATTTGATTAACAATG 108

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec

acute leukemias with rearrangements in the BCR gene first intron: further evidence implicating Alu sequences in the chromosome translocation
Nucleic Acids Res. 17 (19), 7631-7642 (1989)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
14 (sites)
Tatusov,R. and Lipman,D.J.
Using local similarities for pattern detection in nucleotide/protein sequences
Unpublished

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
15 (sites)
Groffen,J., Stephenson,J.R., Heisterkamp,N., de Klein,A., Barten,C.R. and Grosveld,G.
Philadelphia chromosome breakpoints are clustered within a limited region, bcr, on chromosome 22

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
Cell 36 (1), 93-99 (1984)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
16 (bases 37662 to 37903; 49889 to 50062; 50626 to 50719)
Shivelman,E., Lifshitz,B., Gale,R.P. and Canaan,L.E.
Fused transcript of abl and bcr genes in chronic myelogenous leukaemia
Nature 315 (6020), 550-554 (1985)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
17 (bases 43469 to 44077)
Chen,S.J., Chen,Z., d'Aurion,L., Le Coniat,M., Grausz,D. and Berger,R.
Phi+bcx- acute leukemias: implication of Alu sequences in a chromosomal translocation occurring in the new cluster region within the BCR gene
Oncogene 4 (2), 195-202 (1989)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
18 (bases 1 to 84539)
Chisoe,S.L.
Sequence of the human abl and bcr genes
Thesis (1994) University of Oklahoma

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
19 (bases 1 to 84539)
Roe,B.A.
Direct Submission
Submitted (08-MAR-1994) Bruce A. Roe, Dept. of Chem. and Biochem., University of Oklahoma, 620 Parrington Oval, Rm 208, Norman, OK 73019, USA

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
20 (bases 43316 to 43837)
van der Pelz,M.J., Shivji,M.K., Allen,P.B., Heisterkamp,N., Groffen,J. and Wiedemann,L.M.
Nucleotide sequence of both reciprocal translocation junction regions in a patient with Ph positive acute lymphoblastic leukaemia, with a breakpoint within the first intron of the BCR gene
Nucleic Acids Res. 17 (1), 1-10 (1989)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
21 (bases 80055 to 82432)
Andrews,D.F., III., Tompkins,C.K., Hendrickson,S.L. and Singer,J.W.
Cloning and sequencing of the human c-abl 3' untranslated region
Oncogene 5 (3), 441-444 (1990)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
22 (bases 24740 to 25039)
de Klein,A., van Agthoven,T., Groffen,C., Heisterkamp,N., Groffen,J. and Grosveld,G.
Molecular analysis of both translocation products of a Philadelphia-positive CML patient
Nucleic Acids Res. 14 (17), 7071-7082 (1986)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
23 (sites)
Heisterkamp,N., Stam,K., Groffen,J., de Klein,A. and Grosveld,G.
Structural organization of the bcr gene and its role in the Ph' translocation
Nature 315 (6022), 758-761 (1985)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
24 (sites)
Litz,C.E., McClure,J.S., Copenhaver,C.M. and Brunning,R.D.
Duplication of small segments within the major breakpoint cluster region in chronic myelogenous leukemia
Blood 81 (6), 1567-1572 (1993)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
93200517
25 (sites)
Soekarman,D., van Denderen,J., Hoefsloot,L., Morel,M., Meuwissen,T., van Baal,J., Hagemeijer,A. and Grosveld,G.
A novel variant of the bcr-abl fusion product in Philadelphia chromosome-positive acute lymphoblastic leukemia
Leukemia 4 (6), 397-403 (1990)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
90294679
Location/Qualifiers
1..84539
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="9q34"
/map="9q34"
join(U07561.1:29132..35962,U07562.1:1..59012,1..80445)
/gene="ABL"
join(U07561.1:29132..29267,49889..50062,50626..50921,58588..58860,67954..68038,68685..68862,70693..70877,73175..73527,74828..74917,75260..75424,78731..80445)
/gene="ABL"
/note="ABL is the cellular homolog proto-oncogene of Abelson's murine leukemia virus and is associated with the t(9;22) chromosomal translocation with the BCR gene in chronic myelogenous and acute lymphoblastic leukemia; alternative splicing using alternative first exon 1b"
/citation-[8]
/citation-[10]
/citation-[11]
/codon_start=1

Query Match 86.3%; Score 16.4; DB 10; Length 84539;
Best Local Similarity 94.4%; Pred. No. 53;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AAGGATCGAACAATGCC 18
Db 54910 AAGGATCGAACAATGAC 54893

RESULT 3
LOCUS HST66616 4022 bp mRNA PRI 18-SEP-1996
DEFINITION Human SWI/SNF complex 170 KDa subunit (BAF170) mRNA, complete cds.
ACCESSION U66616
NID 91549240
VERSION U66616.1 GI:1549240
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4022)
Wang,W., Xue,Y., Zhou,S., Kuo,A., Cairns,B.R. and Crabtree,G.R.
Diversity and specialization of mammalian SWI/SNF complexes
Genes Dev. 10 (17), 2117-2130 (1996)

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
2 (bases 1 to 4022)
Wang,W., Cote,J., Xue,Y., Zhou,S., Khavari,P.A., Biggar,S.R., Muchardt,C., Kalpana,G.V., Goff,S.P., Yaniv,M., Workman,J.L. and Crabtree,G.R.
Purification and biochemical heterogeneity of the mammalian SWI-SNF complex
EMBO J. 15 (1996) In press

JOURNAL MEDLINE
REFERENCE
AUTHORS
TITLE
3 (bases 1 to 4022)
Wang,W., Xue,Y., Zhou,S. and Crabtree,G.R.
Direct Submission
Submitted (12-AUG-1996) Howard Hughes Medical Institute, Stanford University, Beckman Center B207, Stanford, CA 94305-5428, USA
Location/Qualifiers
1..4022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="Jurkat T cells"

gene 23..3664
/gene="BAF170"
CDS 23..3664
/gene="BAF170"
/note="similar to human BAF155 and yeast SWI3; contains a region similar to the DNA binding domain of myb and a predicted leucine zipper domain; the C-terminus of the protein is highly proline-rich"
/codon_start=1
/product="SWI/SNF complex 170 kDa subunit"
/protein_id="AAC50694.1"
/db_xref="pid:g1549241"
/db_xref="gi:1549241"
/translation="MAVRRKKGDPNKKYIEADDTVTQEDNRLMLCKNKKYIQAEPP
TKRSLSLVYOLQFOEEVFGKHNAPLTKLPICQFLDKAGSGSCIIIAAYKRS
DGMRRYDFONPSRMDNEMETIEKSLVONNLSRPNIFLCEPEIEPLKLNQI
KRIHOGVTDKNNASHVYVPGNLEEEEMVPMKRDQVLLHMGYDSDYDTPA
SEIASVEDAPTPPEKPKVHAKWIIDTDFNEMNEDYEDVDNKKPVRKRSKAT
LIDVNSPDSRDRCKGKTKRKRSPPSPPEYEECKGKSPITPKSKGRHEE
EDEDLTKDEPSVPEVEVLPTKVNKKKSESAPVGGTIDDEDSDMETTG
KDEDSTNGKDEQKNDPLHEDNTEQTHIIPSYAMFYNSVHALERALPEPF
NGKNSKTEIYLAYNEMIDYRLNPOELTSTACRNLACVCAISVHAFLEQMG
LINVYDASRPTMGPPPTSHHVADTSGSLVPLQPTPOOTSQOMLNPDRGK
EAPTMONGELTDMYTKKNAKSKSAATREMTQETLILLETALMYKDMNVS
EIVGSKTQECILHLRLPIEDPILEDSSASGLPAYQPIPSQSGNPMSTYAPLAS
VDPYVAAKASALEESKMEEVPTALVEHVRVEEAAVGTGADPAFGLESIGI
AGTSDPERIESESGDEARVGGQATDEKEREPEEGGALIEEKERTSAPKDE
EKGKEDSESEKSDGDPIDPEKEKEKEGEVLEKVESGKRTKVERDIEG
NSTAAALAAALAAAKHLAAVEERIKSLVALVETOMKLEILRHEELETMD
RREALEYOROOLADROAFHMDLQKPEMRAQOHFOOMHQQOOPPALPGSOP
PERTGAAPRAVHGLAVAPASVPAPAGSAPSGSGSQIQAGSTRPOOQAPGA
PPGAVPVGPPPPGPHGSPPTNQTPTSMGPAVAGSGHPAGANPGLPGMPP
PPPPASIIIPFGSLADISINLPAPNLGSPSPVPPPTLPPLNLPVMANPLHNL
PATTPMPSLPLPGSGSAQSPAIYAAVVOGNLPLSASPLDPPTLPDPPTAPSPG
TVPVPPPG"

BASE COUNT 1084 a 1153 c 1051 g 734 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 11; Length 4022;
Best Local Similarity 94.4%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 AGGATCAGAACATGCT 19
|||||

Db 528 AGGATAAGAACATGCT 545

RESULT 4
H516962R/c 169 bp DNA PRI 19-OCT-1995
DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 169g2,
ACCESSION reverse read cpg169g2.rta.
NID 259694
VERSION 259694.1 GI:1031607
KEYWORDS Cpg Island; genomic MseI fragment.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 169)
AUTHORS Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.
TITLE Direct Submision
JOURNAL Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,
CB10 1RO, England. E-mail contact: humquerry@sanger.ac.uk
2 (bases 1 to 169)
REFERENCE Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.
AUTHORS Purification of Cpg islands using a methylation DNA binding column
JOURNAL Nat. Genet. 6 (3), 236-244 (1994)
MEDLINE 94282070
COMMENT Vector: pGEM-5zf(-)
Clones are available from the UK MRC Human Genome Mapping Project

Resource Centre, Hinxton, Cambridgeshire CB10 1RO, UK. See URL:
http://www.hgmp.mrc.ac.uk/ for details
or contact: biohelp@hgmp.mrc.ac.uk.
Location/Qualifiers
1..169
/organism="Homo sapiens"
/db_xref="taxon:9606"
/sex="male"
/dev_stage="adult"
/tissue_type="blood"
/clone_1lb="CGI-1"
/clone="169g2"

BASE COUNT 41 a 46 c 32 g 47 t 3 others
ORIGIN

Query Match 84.2%; Score 16; DB 9; Length 169;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGATCAGAACATG 16
|||||

Db 108 AAGATCAGAACATG 93

RESULT 5
AC005543 79810 bp DNA PRI 26-FEB-1999
LOCUS Homo sapiens chromosome 4 clone B320D6 map 4q25, complete sequence.
DEFINITION AC005543
ACCESSION 94309859
NID AC005543.2 GI:4309859
KEYWORDS HTG.
VERSION human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 79810)
AUTHORS Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
TITLE Direct Submision
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 79810)
JOURNAL Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submision
JOURNAL Submitted (26-AUG-1998) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
3 (bases 1 to 79810)
REFERENCE Stone,N.E., Schmutz,J.J., Shang,J., Cox,D.R. and Myers,R.M.
AUTHORS Direct Submision
JOURNAL Submitted (26-FEB-1999) Department of Genetics, Stanford Human
Genome Center, 4005 Miranda Avenue, Palo Alto, CA 94304, USA
On Mar 1, 1999 this sequence version replaced g1:4225912.
COMMENT Quality: Phrap Quality >=40 98.9% of Sequence;
Estimated Total Number of Errors is 0.8.
Location/Qualifiers
1..79810
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="B320D6"
/chromosome="4"
/map="4q25"
/clone_1lb="Caltech/Research Genetics Whole-Genome BAC
library"

BASE COUNT 24697 a 15875 c 15934 g 23304 t
ORIGIN

Query Match 84.2%; Score 16; DB 42; Length 79810;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGATCAGAACATG 16
|||||


```

/note="alternate gene name: yula; similar to
scb1c1-6-phosphate 2-dehydrogenase
/codon_start=1
/transl_table=11
/protein_id="CAB15111.1"
/db_xref="PID:el184200"
/db_xref="GI:2635618"
/db_xref="GI:2635618"
/db_xref="SWISS-PROT:P40747"
/translation="MYKH1WDSERAAQLPKGYEELYSRNLIGSDRTVCNMGNGNTSM
KTEKEDRGRELEIMVWKGSDLATMKAHNFSGKLDDIRPLIKRDKQDDEMDVYL
SHCMIDSKRPSIEHLIAPLKYKHVDHPDAIISICCANQGOIADIDYDPAVY
VYVRRGFTLSPKRIAGVANNPRAEYLMEKGIYWGTSSECTYKTSITIOEAOY
INDRIQHEVFGKRTQPLPEDKRKQILGIMFVINGAVEEKMILSTDDHVDLEF
VYSVQAPALSOIGACPDHLVHKRVPILIDMNPETQDVHKLADLRKGEVETSEYQ
AVTRNQDQGDQIFESAPRVILPGIGMNTGYSMSKVSGLYRRAIYAKKATAL
GQFVSLHNESEYHVEYRLEIKLTLAPAEASRKVALITGAGAGISAAORREFAE
GGHVIYADINIEGAOKIAGEINDAVGKGRAMVKNMVTKEPDVQSAFEBAALAYGID
IYVNNAGLATSPPDETSLKENLNNVYGTGYELVRAAFQMKHONGSGMVPFGS
KNSVYAGKASATSVKALETHLARCIAEKGEGFIRVSVLPDAVLQCSALWGSWR
ERRAAVGIPEQLEHYRKRALLVNIYEDIAEIAEFASSKAEKTTGCMITVGG
VPAAFTR"
/complement(6109..8125)
/gene="tlpB"
/complement(6109..6124)
/gene="tlpB"
/complement(6137..8125)
/gene="tlpB"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"
/protein_id="CAB15112.1"
/db_xref="PID:el184201"
/db_xref="PID:92635619"
/db_xref="GI:2635619"
/db_xref="GI:2635619"
/db_xref="SWISS-PROT:P39217"
/translation="MGKFTQWTKQPSISPLIAFLAVLILPVGVLAYFSYOSANAL
DBELISSAKVNEELNSTQNKLEDKVKAIDYSEYVDLILKKRTILKKFROYT
TINDVGATYANSEDKLTKYRPSGVKPGFDPGRWYKQVAEKQAFSEPYDEA
TGDIVATTISKQKDGSGVIALDLNDEVLTASRKIRIGEFATITGNKKYIAHPTI
KPGTQSGDMTQNVYSKKEGSEYFEKGEKMAFTNLTGMKIGATFEVSELDAS
SPVLNTAVIILCVSIYIGILITLYIRATTKPKIYVSAAKISSGDLFEVIDHKN
EFGQGESFNEMSAASRVISYIOTSEVNAASSSELTASAOTSRATSHITLAEQF
SPGNEASKELETSSNHLSONMGDISKVAQASTITKSSIOGSEASGESEKIVERTG
QKRTIDQYQKAEAVKLETSQDITSLYNTINGIADTINLLALNALEAARAGTIG
KGFVYAEVRKLAQSADSAKEIGELDEIVAREITSLSMFSQNVHREKELQIDQ
TAESEQIYEMTQISGEQNLNATVEQLSAGQSEVSAVEISAVAKESAGIODIA
ASAEQOLASMEETISSAETLANMAEELADITKKFKIES"
/complement(8204..8228)
/gene="mcpA"
/complement(8204..10224)
/gene="mcpA"
/complement(8239..10224)
/gene="mcpA"
/function="chemotaxis deficiency toward glucose and
alpha-methyl-glucoside"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemotaxis protein"

```

```

Query Match      83.2% Score 15.8; DB 1; Length 217420;
Best Local Similarity 89.5%; Pred. No. 1.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AAGGATCAGACAACTGCT 19
||||| ||| ||||| |||
Db 133918 AAGGACAGCAATGCT 133936

```

```

RESULT 7
LOCUS AB005238 75188 bp DNA PLN 05-FEB-1999
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MKP11,

```

```

ACCESSION complete sequence.
AB005238
NID 92264310
VERSION AB005238.1 GI:2264310
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
clone:MKP11.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (sites)
REFERENCE
AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
Miyajima,N. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned P1 clones
JOURNAL DNA Res. 4 (3), 215-230 (1997)
97471969
2 (bases 1 to 75188)
REFERENCE
AUTHORS Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2, 1532-3, Yana, Kisarazu, Chiba 292, Japan
(E-mail:ynakamu@kazusa.or.jp, Tel:+81-438-52-3935,
Fax:+81-438-52-3934)
FEATURES
source
1..75188
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MKP11"
/clone_lib="Mitsui P1"
BASE COUNT 24812 a 12842 c 13514 g 24020 t
ORIGIN

```

```

Query Match      83.2% Score 15.8; DB 7; Length 75188;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AAGGATCAGACAACTGCT 19
||||| ||| ||||| |||
Db 33173 AAGGACAGCAATGCT 33191

```

```

RESULT 8
AB005246/C 77999 bp DNA PLN 05-FEB-1999
LOCUS AB005246 Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone: MUP24,
DEFINITION complete sequence.
ACCESSION AB005246
NID 92264318
VERSION AB005246.1 GI:2264318
KEYWORDS HTG.
SOURCE Arabidopsis thaliana (strain:Columbia) DNA, clone:lib:Mitsui P1
clone:MUP24.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (sites)
REFERENCE
AUTHORS Sato,S., Kotani,H., Nakamura,Y., Kaneko,T., Asamizu,E., Fukami,M.,
Miyajima,N. and Tabata,S.
TITLE Structural analysis of Arabidopsis thaliana chromosome 5. I.
Sequence features of the 1.6 Mb regions covered by twenty
physically assigned P1 clones
JOURNAL DNA Res. 4 (3), 215-230 (1997)
97471969

```

REFERENCE 2 (bases 1 to 77999)
AUTHORS Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (02-JUL-1997) to the DDBJ/EMBL/GenBank databases.
Yasukazu Nakamura, Kazusa DNA Research Institute, Laboratory of
Gene Structure 2: 1532-3, Yama, Kisarazu, Chiba 292, Japan
(E-mail: ynakamura@kazusa.or.jp, Tel: +81-438-52-3935,
Fax: +81-438-52-3934)

FEATURES
source
1. .77999
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone="MOP24"
/clone_id="Mitsui P1"

BASE COUNT 24666 a 13395 c 14042 g 25896 t
ORIGIN

Query Match 83.2%, Score 15.8; DB 7; Length 77999;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCAGACAAATGCT 19
|||||
Db 12726 AAGGATCAGACAAATTTCT 12708

RESULT 9
AF049236/c
LOCUS AF049236 60000 bp DNA PLN 22-APR-1998
DEFINITION Arabidopsis thaliana putative transmembrane protein gip (Atg1),
putative nuclear DNA-binding protein G2P (Atg2), E1 protein
(ATEM1), putative chlorophyll synthetase (Atg4), putative
transmembrane protein G5P (Atg5), putative acyl-coA dehydrogenase
(Atg6), and calcium dependent protein kinase genes, complete cds;
and unknown genes.

ACCESSION AF049236
NID AF049236
VERSION g3068702
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 16488 to 21370)
Gaubier, P., Wu, H.-J., Laudie, M., Delseny, M. and Grelliet, F.
A chlorophyll synthetase gene from Arabidopsis thaliana
Mol. Gen. Genet. 249 (1), 58-64 (1995)
96140448
2 (bases 1 to 60000)
Gaubier, P., Wu, H.-J., Berger, C., Cooke, R., Delseny, M. and
Grelliet, F.
Sequence of a 60kb contig on chromosome 3, the ATEM1 locus
unpublished
3 (bases 1 to 60000)
Grelliet, F.
Direct Submission
Submitted (19-FEB-1998) LDBMP, Universite de Perpignan, 52 Avenue
de Villeneuve, Perpignan 66860, France

FEATURES
source
1. .60000
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/complement(join(922..1362,1541..>1786))
/product="hypothetical protein"
/complement(join(922..1362,1541..1786))
/note="hypothetical protein; ORF1: similar to sequence in
Genbank Accession Number AC002330"
/codon_start=1.

gene
mRNA
CDS
/product="unknown"
/protein_id="AAC14403.1"
/db_xref="PID:g3068703"
/db_xref="GI:3068703"
/translation="MILYLVEMMDINRSRLNENPFLSTRHRYOQKLPQGFPA
SVPRCSAGCMKPNSSHLKNDASEFPRLKPEGRKEFPHQNDQAVYLSQTQI
DMVLPKQDSVSESDDRSSVGGDSNGSLTSESTERSDMSPTIIEFPLDAALAV
TSASQRRKRLSTLSETVRSQCFSPKAGLHVLPLWSTKHKRCGVKNMFSPSVH
LQPFIRKDN"
complement(join(<2709..2935,3332..3641,3930..4421,
4329..>6376))
/product="hypothetical protein"
complement(join(2709..2935,3332..3641,3930..4421,
4329..6376))
/note="hypothetical protein; ORF2: similar to sequence in
Genbank Accession Number AC002330"
/codon_start=1
/product="unknown"
/protein_id="AAC14404.1"
/db_xref="PID:g3068704"
/db_xref="GI:3068704"
/translation="MCHNLFYMRSLKLAGCKGTQYALNPSAPTPPPGNSGTG
GGGGSGGTGGVGDKLQHLSDHLRYNSVRSKSRTPPTOPNAVSPFLLPCL
PYDLELPQIDPCLKPYDIYEKNAQVRRYRLENCQEFKSGAYLEQCAIFRGISPKLF
RSLRSRQAHVDYNAKAVLASWLSGGDELIGTISMDCCGRNLECPKATLVSGYD
ESVYDPCVSGASSEMNEDECTSOEVIYDMSFCIGDEVCVRVKRLASRFA
MLYGFREMKRATINFTONGISVEGMRAAEIFSTNTLDNFPVNVLELKLARFCA
DEKSAQDSHLAHLVNSLDEAMLLIEYGEAEAYLVAAQLOVLRLLPSMHPNYI
KIPCSAEGRERLASLGHSFTLYPFLSLQIEMEDMKNTYMLERLVECAVDSMEQ
LAYHQLGVVLEREYKQDQRMFNAAYEAHLYSLVAVARTPKRDRYSAYKINSL
ISDRKATGMHQRNSLYSCGKREKLLDIDIERPDTLTFPKRAVALVENQGAAL
AELNRKILGFASPOCLEMRAMVTSIGMDYEGALNDIALLTLEENFMKWKHGDH
VELRLPAAQMSQDCWMOLOYDRWSVDDIGSLAVVHMLANDPGKILFRSLL
RLNCKAKAMSLRLARNHKSSEHRLVYEGMIIYDTGHEEALAKAESIORSFA
FLKAYVALSTDPDSSNVVIOLGALCPSPGKGLGALLNGSVYDCELDIA
ADCYTNLTIKRPAHOGIARVYLNKORRAADDEMRTLEKANNASAYEKREYCD
REKASDLCIATQIDPRLTPYRRAVALDQDKRESAIDELSLAISFKPDQLHLHR
AAFYDSMGESASIKDEEALCIDPGADVLELYHKAREPDKR"
complement(join(9320..9605,9694..9840,9918..10174,
10350..10662))
/product="hypothetical protein"
complement(join(9408..9605,9694..9840,9918..10174,
10350..10537))
/note="hypothetical protein; ORF3: similar to Arabidopsis
thaliana clone 240M377 sequence found in dbEST Accession
Number N65910"
/codon_start=1
/product="unknown"
/protein_id="AAC14405.1"
/db_xref="PID:g3068705"
/db_xref="GI:3068705"
/translation="MMNSTEESSEWEYRPGMLVQRDDTASSDHRFLDPDSANAF
AOTRITVSHSSHHDLHSIAHATFGDKVKRAKQCTGLSESLKILFRGVRDQADL
OAGVRYASLTVYVEDTNKRVEDQPPVYTKEMKAAIAYNAATVEYDKLSDRYALE
VAVNGTQVAVREFDMAEILMRQLLDLDIEAGDAKAVKRAVEVRIDQLQEVAVLE
KARCSNPFVDSKAAAVSTEWESGNGVSLNPPPSAPSNVYQDWEKPD"
complement(12025..13047)
/gene="Atg1"
complement(<12272..>13042)
/gene="Atg1"
/product="putative transmembrane protein gip"
complement(12272..13042)
/gene="Atg1"
/note="similar to Bradyrhizobium japonicum CycD: SwissProt
Accession Number P45401"
/codon_start=1
/product="putative transmembrane protein gip"
/protein_id="AAC14406.1"
/db_xref="PID:g3068706"
/db_xref="GI:3068706"
/translation="MARLLFRSSQILRSIQRNPQISSSESPDPIFSLTASPD
PSRLSSITFLRSLSIARGPTRRKIDIGAKARQMDQRRLMTALTITSCIAGFYIYL
NPFQDLVFTLTPSDAKERFENPTKPKFRIGLVLEGSVAQPAQSEMEFVITDLIT

Query Match 83.2%; Score 15.8; DB 8; Length 1348;
Best Local Similarity 89.5%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGGATCAGACATGCCCT 19
|||||
DB 256 AAGGATCAGATGCCCT 274

RESULT 11
ATAC006223
LOCUS ATAC006223 103194 bp DNA PLN 20-FEB-1999
DEFINITION Arabidopsis thaliana chromosome II BAC F22D22 genomic sequence,
complete sequence.
AC006223
AC006223
AC006223.2 GI:4263694
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; euroids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 103194)
Lin.X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., Vanaken,S.E.,
Barstead,M.E., Mason,T.M., Bowman,C.L., Rongning,C.M., Benito,M.,
Carrere,A.J., Greasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
Fraser,C.M. and Venter,J.C.
Arabidopsis thaliana chromosome II BAC F22D22 genomic sequence
Unpublished
2 (bases 1 to 103194)
Lin.X. and Kaul,S.
Direct Submission
Submitted (15-DEC-1998) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA, xlinetlgr.org
3 (bases 1 to 103194)
Lin.X.
Direct Submission
Submitted (20-FEB-1999) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Feb 24, 1999 this sequence version replaced gi:4176358.
Address all correspondence to:
Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr.
Rockville, MD 20850, USA
e-mail: xlinetlgr.org
BAC clone F22D22 is from Arabidopsis chromosome II and is near the
molecular marker TEm5.
The orientation of the sequence is from SP6 to T7 end of the BAC
clone.
Genes were identified by a combination of three methods: Gene
prediction programs including GRAFT (available by anonymous ftp
from arthur.epm.ornl.gov), GeneFinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://genomic.stanford.edu/~chris/GENSCANW.html), and NeuplantGene
(http://www.cbs.dtu.dk/netgene/cbsnetgene.html), searches of the
EST database at TIGR (http://www.tlgr.org/tld/at/ac.html).
Annotated genes are named to indicate the level of evidence for
their annotation. Genes with similarity to other proteins are named
after the database hits. Genes without significant peptide
similarity but with EST similarity are named as 'unknown' proteins.
Genes without protein or EST similarity, that are predicted by more
than two gene prediction programs over most of their length are
annotated as 'hypothetical' proteins. Genes encoding tRNAs are
predicted by tRNAscan-SE (Sean Eddy,
http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
identified by RepeatMasker (Arian Smit,
http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of

FEATURES

genomic sequence that are not annotated as genes but have predicted
exons by GRAFT are annotated as misc features.

Location/Qualifiers
1..103194

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/chromosome="II"

/map="TEm5"

/clone="F22D22"

1..1596

/note="overlap with BAC clone T32F6
(AC005700:98829..100424)."

join(80..133,321..3476,3704..3808,3904..4059,4278..4376,
4557..>4796)

/gene="F22D22.1"

<80..>4796

/gene="F22D22.1"

/note="132F6.24"

join(80..133,321..3476,3704..3808,3904..4059,4278..4376,
4557..4796)

/gene="F22D22.1"

/codon_start=1

/product="putative myosin II heavy chain"

/protein_id="A015381.1"

/db_xref="pid:94263695"

/db_xref="gi:4263695"

/translation="MEETQVTSSEVPYKGDROCRKCLDLQAVNGEVPKEKEE
DGEFTKVEKADDAKADHVEVEKKEKESGSGOSELSEGEKKELEL
EVNGELKRYSENPDLDELAKKELEPEKKEKESGSGOSELSEGEKKELEL
OLKSLDELQSHDADKDETEKKEFDLGLTESSRKRLTELEGLRSAPKAPKE
ELHKRSASHADESOALEFSELKSTESKAKEMEKASLOQELKEKNEKSENEK
EALKSASGELAAVOEELASRLTEOVYSSTELALIDELTQDEKRLTANKEA
ELSYVDLDAQTKGLQAKISEQKINSKLAELKELELSKQDEKRLTANKEA
EVLKKEALENVAEVTNVAETVEVCEKLEPEKTESPENSRTDALSQANSSEL
EOKLSLELSESGSAAATOKRLTEEDVYRSQAAFEKSOIKLEKRTAEO
KNALEQDLNLQKSSAPREKELSEKSSSELQRAIEVABEKKAQATQOETQKA
SELESLTQSSSARNSELEDLALQKAHEHEDRANTTQRTLEGLQSSQSHED
AEGRLKDELILQTERKRIQLELEVSSLEKKGTEADSVKGLQVAVELSTLAFQ
VSSSLLEALNATENEKELTENLNAVSEKKEKATVDEYSKISEENLESLTNE
LNTQGLESTIENDSKAGLOSEVEMKSKKEELDEKGRIDEATTKRMLELHQ
SLSDSEHRLKAMEEPTSRDSEASLTETKRLDLEKTKSVBEOLAEASGSSLEK
LEQTKGRILAAEVSVEKIKQEPDQAKESLOSSSESLAETNNLKKIKIEBELIG
SGSVKEKRLRLEKLEIFENOKETESSDLVEKLTENQLEIKLKHESGAVDTR
KVELDALSKLNKESTIEELGAKQGESGDAEYNLKNLELANHGSANLQF
KLSALEAEKQDANLELASKTTIEDTLTQLESSEKLSQIEKLAVAAYESVLESH
ELEKTLSEVNAKOLENENNAATASVKAELTSLQOEHEHAGEEDVLEQVLOKRE
LOAASSTDEOKRLKRLKRLSLINLNMOKTIINRCKSSOMLKRLETEAMDGVSRDI
DISPSSPTKRKSKKPEASLSSSSSSGNVTPPTOTASTSHMTATVITVALISVIIG
ITLGRKY"

5226..5915

/note="exon predicted by xgrail, quality excellent"

complement(5283..5553)

/note="exon predicted by xgrail, quality
excellent_shadowexon"

complement(5638..5906)

/note="exon predicted by xgrail, quality
excellent_shadowexon"

6067..6180

/note="exon predicted by xgrail, quality excellent"

complement(6356..6536)

/note="exon predicted by xgrail, quality good_shadowexon"

6423..6587

/note="exon predicted by xgrail, quality excellent"

join(<7175..7269,7839..8808,8890..>9075)

/gene="F22D22.2"

<7175..>9075

/gene="F22D22.2"

/note="predicted by genefinder"

join(7175..7269,7839..8808,8890..9075)

/gene="F22D22.2"

/note="hypothetical protein"

/codon_start=1

QY ... 1 AAGGATCAGAACAAATGCT 19

|||||
AAGAATCATACCAATGCCCT

YSCCL8543 37730 bp . DNA PLN 01-AUG-1997
Saccharomyces cerevisiae chromosome XII cosmid 8543.

g2258165
U20618.1 GI:2258165

baker's yeast strain-S288C (AB972)

Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales
Saccharomycetaceae; Saccharomyces.

1 (bases 1 to 37730)
Johnston, M., Hillier, L., Riles, L., Albermann, K., Andre, B.,

Ansong, W., Benes, V., Bruckner, M., Dellus, H., Dubois, E.,
 Duermasthoff, A., Enlihan, K.D., Floeh, M., Goffeau, A., Hebling, U.,
 Heumann, K., Hauss-Neitrl, D., Hilbert, H., Hilger, F., Kline, K.,
 Kotter, P., Louis, E.J., Messing, F., Mewes, H.W., Miosga, T.,
 Mostl, D., Müller-Auer, S., Neutwig, U., Obermann, B., Piravandi, E.,
 Pohl, T.M., Portetelle, D., Punelle, B., Reichmann, S., Rieger, M.,
 Rinke, M., Rose, M., Scharfe, M., Scheens, B., Scholler, P.,
 Schwagerl, C., Schwarz, S., Underwood, A.P., Uristatanz, L.A.,
 Vandenbol, M., Verhasselt, P., Vierendeels, F., Voet, M., Volckaert, G.,
 Voss, R., Mamuth, R., Medler, E., Medler, H., Zimmermann, F.K.,
 Zollner, A., Rauli, J. and Hohelsel, J.D.
 The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIII
 Nature 387 (6632 Suppl.), 87-90 (1997)

The sequence of *S. cerevisiae* cosmid 8543

Unpublished (1994)
3 (bases 1 to 37730)

Waterston, R.
Direct Submission
Submitted: 03-FEB-1995

Submitted (03-FEB-2004) (bases 1 to 37)

Cherry, J.M.
Direct Submission
Submitted 116-JUL

Submitted (10-2001)
On Jul 16, 1997 to
Submit

Depart
Genom
Cancer

St. L
e-mal

NEIGHBORING COSMOS

3

This sequence Inc overlaps with the

nucleotides, and nucleotides.

Location
1. .3773
Jordan's

```

/organism
/db_xref
/map="XI

```

```

/imap- v1
/strain-
compleme

```

```
complete
/gene="C
complete
```

```

/ene="C
/ote="Y

```

/codon_s
/product

/protein_id="AAB64528.1"
 /db_xref="pid:g2258166"
 /db_xref="gi:2258166"
 /translation="MSDTNTSIPNTSSAREAGNACQPSISSSSNTSTTTNTSSAS
 LSSPSTSELSTIRPIVAAVDFNPILKKSSSLSSVOQETIYLKNSGMD
 GLVDSNGKVRNGWFPONGRPLROSHLRKSSPMKYSKSSRRSSLSKNSAY
 LHPRPSKRRGSSSTLSASLNHNATSSGHNNTYMANNSPSPANDASHITPOS
 NENSASLSODMTKRSADGSEEMNTNAINNNETNLOYSGERAGPPLAETIKLPLE
 ELEMINGISNASTWSPILITKTSYDRKLVYKNDLDIYCSPLISNIMESDI
 CDEKPEFPNDHLVNLTRDLRKNANIEDSTRSQSEQONRSLMERKODSETDG
 NNSINDDNHNNKNEEENAGSSSLNSAPDLQNSRYVAAPSRSSILKASDI
 YHRSKIKTEITDLVYTTKTAHKMFLKENRNFYKYPDLISDIIVPOLGCRMG
 HEIKASCSKEIKIKIFKGLSSISRSISNHLTFDQAPHRKMDTANDKNDENNAC
 SRTEGDGKLEIVSDHDLVSPJSGKRVSTSTDTITPKRSSFTVENDEMENEVL
 GPRNSVNSVTPRTSIQNSTLEDFSPSNKPKSAKSIYEMADVEE"
 /complement(2279..2626)
 /gene="YLR311C"
 /complement(2279..2626)
 /gene="YLR311C"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr311cp"
 /protein_id="AAB64518.1"
 /db_xref="pid:g662134"
 /db_xref="gi:662134"
 /translation="MKLTREKNDCLVGVSTIPPLNFTLTLFLALIEVHSLSS
 LSLRIFYHNVCYPSLFLFCFVLPFFYSVRILLYLILRSFYELNILLYGAEN
 SRQSPGYVIR"
 /complement(2651..3847)
 /gene="YLR312C"
 /complement(2651..3847)
 /gene="YLR312C"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr312cp"
 /protein_id="AAB64517.1"
 /db_xref="pid:g662133"
 /db_xref="gi:662133"
 /translation="MSEEDHNLVLRILRKGRGEGSSKSEISLDSLHSSPAGE
 DDEDFDADVLNNTSSESAQMNRIYDRTSESNAGVNIDQGVPTISSEFDLSSS
 NNGGTVPJPMESGSLKNDSTIRNSLSDHILDKSEGSAKLMHVIYMLSSLSMTS
 YLALEYSLTGDLVAGFKSOOSLRNNERKLLYGNIDFYDKSSDSSLSOMAPSGY
 YVDEDNHAIAPLADDDLMGRYKRTDYLITMYTKAMKDMGRIRIKINGRITKEL
 FLNSTRSAQESLRYLHKEOKRRRKRLLFVLLHNTYRQPSPIKRTYFHSCKAKOCS
 GSRLOQLKLRKSKNKPFRVFOVKRQTNMFVQLKRRFGKLQHSRMYKAMSECRKN
 YFKCKH"
 /complement(4494..5255)
 /gene="MRPL15"
 /complement(4494..5255)
 /gene="MRPL15"
 /note="Ylr312w-ap"
 /codon_start=1
 /product="Mrp15p: mitochondrial ribosomal protein yml15"
 /protein_id="AAB64533.1"
 /db_xref="pid:g2258171"
 /db_xref="gi:2258171"
 /translation="MENSMAFISRLRPYALNCNQSVRTVYLKGRINGLRD
 PESYLRNPSCVLFTEVNAKECODVRSIILDPKYGIMLSNELLOCTLHKSFAFGSP
 YNEKLNLAQAFKLOCTCISLKNKSPAESCEGLOLSNLTQKPAKELTAKNTAC
 TFWKHLNLPETFKMRDPKIDGHINGETTFASVLAFAIGALISTGSKAAAFIIG
 SLDKEDHLSLVNANENVASAKAKISDKENKALV"
 /complement(5764..7356)
 /gene="YLR313C"
 /complement(5764..7356)
 /gene="YLR313C"
 /note="Note that a frameshift between nucleotides
 5765-5805 would extend this gene. However, no mistake in
 the sequence is apparent"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr313cp"
 /protein_id="AAB64516.1"
 /db_xref="pid:g662132"

/db_xref="gi:662132"
 /translation="MISSELTNDQIIDLISDYKFRRIIASVPEDDRNRONRKN
 KDLTKLSNVQWOLTTDVNDLMLRKRLDSCADSLNDLDRKAKOSKLSRLDAKFH
 KLILDIETBEIERNLHLDIGTNNNGDEEDLDFNLDITFESEKINDMDSNGIIS
 IEVERELKTOFTLYFQONTLRIDPDTTTRPLILEETIKIKILGIDLPLVYSVL
 OSSLENTVYLKKAASHAITSTYPTLYGLIDIRIVAOSSISVIFAFCAQIAIVK
 STSRDLSRNEGSLDLEAKMKLITKRYKNKNGDISLDSGSTANSPSSGCP
 ITRKSDPVVAVASPSISITKSSSISRESGVRKNNTSGETNLASVSLAKNTKSSRT
 SEPSRGLGLKAVNSRSPSPQGNTPILGKFRQDQASPPKRVITRPVATKASPA
 NIPPAVDVLSPTVTKRFRERVOKEFAPNSGLRLISTEENLNNSDVSTHNANI
 NNIVFEVSSSNVLPKAGRAF"
 /complement(7589..9151)
 /gene="CDC3"
 /db_xref="sgd:S0004306"
 /complement(7589..9151)
 /gene="CDC3"
 /note="YLR314C: cell division cycle protein"
 /codon_start=1
 /db_xref="sgd:S0004306"
 /product="Cdc3p: component of 10 nm filaments of
 mother-bud neck"
 /protein_id="AAB64515.1"
 /db_xref="pid:g662131"
 /db_xref="gi:662131"
 /translation="MSLKEQVSTIKODPEEERHOQFNDOVKIOWESODHDGYSOYT
 NGTNDSEKFEAEASDVKVEPLGKGTSSQSGQVLPDQPEIKTRQINQYGF
 ANLPKQWHRRSINGEFNLGVGPDIGITTLTKLTFNDDILEANLVKQYEEELAD
 OEEBEGGEGHENQSEQRKRVKIKSYSEYIENGVLNINVIDTEGDFGLNNDKS
 MDPLIKEDSRPOYLDAENKINHSINDRIHACLYFIEPTGHLPLDKFQVSY
 EKNLILPVIAKSDILDEILTSFKRTIMNLIQSNIEFLKPYISNDANSHLSEEL
 FSSLPRVVISNDIVENYSGNVRGSRPGVGLYENDNSPRLKNTLIRKPMETL
 KERTSKLTENYNSKLALGIRKODNVFEFEDPISQLEKTLHEKTLKAKLEEMKT
 VFOQVSEKERRKLQKSETLFAHREKREKLTQLKALEKROKLELINSASPVNKH
 SPVPTKKKGFIR"
 /complement(9822..10283)
 /gene="YLR315W"
 /complement(9822..10283)
 /gene="YLR315W"
 /codon_start=1
 /evidence=not_experimental
 /product="Ylr315wp"
 /protein_id="AAB64519.1"
 /db_xref="pid:g662135"
 /db_xref="gi:662135"
 /translation="MNSQQLHMYVSDLLTTLLISFQEFKQLOQSYTSDEQLOHWE
 LQARDARVSELEARIKQFFITLRSLLFLFEQQLSHLSLTLIDALYKINDLQ
 RLQILDVAIOERTSELAEPENWVRSPSAGDNAIPGLLIQSYINLEEN"
 /complement(10280..10771)
 /gene="YLR316C"
 /complement(10280..10771)
 /gene="YLR316C"
 /codon_start=1
 /product="Ylr316cp"
 /protein_id="AAB64529.1"
 /db_xref="pid:g2258167"
 /db_xref="gi:2258167"
 /translation="MOEVRNELRSLTSYKMATAGKQPMVSVFDPSPKDKRYVAE
 DGRNCENSEPLIDHVMGIRAVGRLEGVDEPANSYLCIDYVYLLHEGCSMSMAL
 IHSVSRVFLTEMQRIGSLKINSGDGYCMNDNKQINSTYEAFQWIGEEYVPGQVDRD
 VCC"

Query Match 83.2%; Score 15.8; DB 8; Length 37730;
 Best Local Similarity 89.5%; Pred. No. 1.1e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 AAGGATCAGACATGCT 19
 DB 10864 AAGGAATGAAACATGCT 10846
 RESULT 13
 HS426N21/c
 LOCUS HS426N21 44826 bp DNA PRI 08-DEC-1998

DEFINITION Human DNA sequence from clone 426N21 on chromosome X Contains EST, STS, GSS, complete sequence.

ACCESSION 282208

NID 93980421

VERSION 282208.1 GI:3980421

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 44826)

REFERENCE 1 (bases 1 to 44826)

GRAHAM,D.

DIRECT SUBMISSION Submitted (07-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk

On Dec 8, 1998 this sequence version replaced gi:1772964.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

IMPORTANT: This sequence is not the entire insert of clone 426N21. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

The true right end of clone d537K23 is at 100 in this sequence. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome X, constructed by the Sanger Centre Chromosome X Mapping Group. Further information can be found at <http://www.sanger.ac.uk/RGP/CHRX>

426N21 is from the library RPI3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://dacpac.med.buffalo.edu/VECTOR>: pCYPAC2.

FEATURES

location/Qualifiers

1. 44826

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="X"

/clone="426N21"

/clone_1lb="RPI3"

39. 350

repeat_region

/note="AluX repeat: matches 1. 312 of consensus"

misc_feature

<1386. .>1723

/note="match: GSS AQ131901"

2516. 2557

/note="MER5A repeat: matches 146. 189 of consensus"

2556. 2641

/note="MIR repeat: matches 28. 116 of consensus"

2642. 2944

/note="AluY repeat: matches 4. 297 of consensus"

2954. 3268

/note="AluX repeat: matches 1. 312 of consensus"

3269. 3282

/note="MIR repeat: matches 124. 137 of consensus"

3283. 3587

/note="AluX repeat: matches 1. 305 of consensus"

3588. 3720

/note="MIR repeat: matches 137. 262 of consensus"

3733. 4032

/note="AluY repeat: matches 1. 301 of consensus"

4425. 4605

/note="AluSg/x repeat: matches 131. 311 of consensus"

4606. 4921

/note="AluSg repeat: matches 1. 313 of consensus"

4922. 4958

repeat_region

/note="AluSg/x repeat: matches 95. 131 of consensus"

4961. 5008

/note="AluSg/x repeat: matches 129. 185 of consensus"

5009. 5142

/note="AluO/FLAM repeat: matches 2. 185 of consensus"

5151. 5464

/note="AluSg repeat: matches 2. 307 of consensus"

5467. 5656

/note="AluY repeat: matches 119. 305 of consensus"

5657. 5969

/note="AluX repeat: matches 1. 299 of consensus"

6219. 6357

/note="AluX repeat: matches 1. 141 of consensus"

6906. 7204

/note="AluSg repeat: matches 1. 298 of consensus"

7288. 7587

/note="AluSg repeat: matches 1. 309 of consensus"

7595. 7694

/note="MIR repeat: matches 107. 223 of consensus"

8309. 8807

/note="AluYB repeat: matches 1. 305 of consensus"

8939. 9241

/note="AluX repeat: matches 1. 304 of consensus"

10115. 10418

/note="AluO repeat: matches 1. 306 of consensus"

11014. 11319

/note="AluSx repeat: matches 1. 308 of consensus"

11483. 11512

/note="15 copies 2 mer at 87% conserved"

12044. 12121

/note="MER5A repeat: matches 85. 162 of consensus"

12548. 12577

/note="15 copies 2 mer tt 87% conserved"

12800. 13093

/note="AluB repeat: matches 17. 311 of consensus"

13197. 13475

/note="AluX repeat: matches 1. 281 of consensus"

13548. 13839

/note="AluSg repeat: matches 1. 300 of consensus"

13979. 14371

/note="LTR37A repeat: matches 7. 426 of consensus"

14397. 14440

/note="22 copies 2 mer tg 89% conserved"

14498. 14579

/note="L2 repeat: matches 2635. 2710 of consensus"

14580. 14890

/note="AluSg repeat: matches 1. 313 of consensus"

14891. 16029

/note="L2 repeat: matches 1456. 2635 of consensus"

16038. 16091

/note="18 copies 3 mer gag 74% conserved"

16570. 16861

/note="L2 repeat: matches 2288. 2517 of consensus"

16862. 17172

/note="AluSx repeat: matches 1. 313 of consensus"

17173. 17369

/note="L2 repeat: matches 2517. 2731 of consensus"

17744. 18046

/note="AluSx repeat: matches 1. 303 of consensus"

19119. 19431

/note="AluSx repeat: matches 1. 299 of consensus"

19861. 20160

/note="AluY repeat: matches 1. 301 of consensus"

20205. 20371

/note="AluO repeat: matches 1. 169 of consensus"

21016. 21331

/note="AluSx repeat: matches 1. 312 of consensus"

21318. 21413

/note="Single clone region. single clone in alu"

21429. 21721

/note="AluSg repeat: matches 1. 295 of consensus"

21767. 22064

/note="AluSg repeat: matches 1. 299 of consensus"

SOURCE Human, cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE 1 (bases 1 to 4626; 889 to 4626)
Tomkinson, B. and Jonsson, A.-K.
Characterization of cDNA for human tripeptidyl peptidase II: The
N-terminal part of the enzyme is similar to subtilisin
Biochemistry 30, 168-174 (1991)
JOURNAL 91105077
MEDLINE 2 (bases 1 to 4626; 1 to 888)
REFERENCE Tomkinson, B.
AUTHORS Nucleotide sequence of cDNA covering the N-terminus of human
tripeptidyl peptidase II
JOURNAL Biomed. Biochim. Acta 50, 727-729 (1991)
MEDLINE 92198394
FEATURES
location/Qualifiers
1..4626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="SPL"
/cell_type="B-lymphocyte"
/tissue="lb="lambda gt10"
24..3773
/codon_start=1
/product="tripeptidyl peptidase II"
/protein_id="AAA56760.1"
/db_xref="GI:339880"
/db_xref="PI:9339880"
/translation="MATATEPEPEPFGHLLPKKETGASFLCRPEYDGRGLVAVLD
TVDPGAPGMQVITDCKPKIVDITDTSGDVNTATPEVDEIGVLSGRVLKIPAS
WTPSGKTHGIGKNGIDYPRKALKERIQKREKIMDPVHRVALAECKROEEDVAN
NGSSQAKLLKEELQSOVELNSFEKYSIDPGVYDCLVHDEWVACIDSNEGDL
SKSTVLTNRYEAGQYSGFTGAEMLNTSVNYDDNMLISTVSGAHGTVAASIAAGH
PEEPERNVAPGAQIISIKIGDRLSTMETGILIRAMIEYINHKCDLVNYSYGEATH
WPNSGRICEVINEVMKNTIIVYSAGNPGCLSTVCGPGGTSSVGVGAVYSPDM
VAEYSLREKLPANQYTWSSSRGSPADGALGVISAPGALASVPNMILRGOLMNGTSM
SSPNACGGAILIISGLKANNIDTVHSVRALENTAYKADNIEVPAQGGITIOYDAY
DYLVQNTSFANKLGFVTVGNNGKITLRDPVOVAAPSDHGVGLEVPENTENSEKIS
LQHLALTSNSWVOCPSHLELNOCRHINIRVDPRLRGLEHYTEVCGDIASPNAG
PLFRPVITAVIAAKVNESHYDLAFTDHFPGQIRRHFEVEGATMAEVYSCSS
EVSARFVIAVOKOVARSHHEFYKSLPEKGTLEAFVVGKALIEFCIARWMS
LSDVNIDYTSFPHGIVCTAPOLINHASEGIRPDVOSLKYEDLAPCITLKNVOTLR
PVSAKTRPLGSRVLPNNRQLYEMVLITNFKRSGEYTPSCPLLCCLLESFDSOL
WIFDOKRQWGSQDAIPHQYSLKEGDTIRIQIRHEQISDLERKDLPTVSRL
SNTLSLDIHENHSFALGKRKSNLTPRYNQPFVYSLPDKIPKAGPGCYLAGS
LTLSTELGKADAVIPVHYLLIPPTKNGSKDKESKEDLKEEFTALRDLKI
QMTKLDSSDIYNELKETYPNYLPLVYARLHOLDAREKRLNEIYDANAVYSHD
QTLAVYIAKMTDPRPDATIKNDMDOKSTLVDAICRGKALADHILHTQADGALS
TDAEGKEEGESPLDSLAETFTETTKTDLFDNKVULTEFAIKHALVNMKMGRLKFAIK
LVEKPTKEMKNCIQLMKLLGWTCHCSFTENWLPIMYPDYCVF"

BASE COUNT 1452 a 885 c 983 g 1306 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 9; length 4626;
Best Local Similarity 89.5%; Pred. No. 85;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCCT 19
|||||||
DB 1412 AAGATCAGGCGCATGCCT 1394

Search completed: September 13, 1999, 15:55:53
Job time: 4581 sec


```

FT misc_feature 11721..11722
FT /tag= h
FT /label= Gap 4.
FT /note= "Bases are estimated to be missing between
FT these two positions in this sequence."
FT exon 13354..13366
FT /tag= 1
FT /label= Exon 5.
FT exon 15044..15160
FT /tag= 1
FT /label= Exon 6.
FT exon 17245..17410
FT /tag= k
FT /label= Exon 7.

PN US5266683-A.
PD 30-NOV-1993.
PE 08-APR-1988; 179406.
PR 15-AUG-1988; US-179406.
PR 23-FEB-1989; US-232630.
PR 17-OCT-1989; US-422613.
PR 22-FEB-1990; US-483913.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 18-OCT-1990; US-600024.
PA 04-DEC-1990; US-621849.
PA 04-DEC-1990; US-621988.
PA 22-FEB-1991; US-660162.
PA 20-DEC-1991; US-810560.
PA 28-JAN-1992; US-827052.
PA 21-FEB-1992; US-841646.
PI Kuberessampath T, Oppermann H, Ozkaynak E, Pang RHJ,
PI Rueger DC;
PI WPI: 93-395405/49.
DR P-SDS; R44746.
DR New pure mammalian osteogenic proteins - induce cartilage and
PT endochondral bone formation when in association with a matrix
PT Disclousure; Columns 73-88; 12pp; English.
CC The osteogenic protein when in association with a matrix can induce
CC at the locus of an implant the full development cascade of
CC endochondral bone formation including vascularisation.
CC mineralisation and bone marrow differentiation. The osteogenic
CC protein can also be used to repair both bone and cartilage in the
CC treatment of osteoarthritis.
SO Sequence 17410 BP; 3833 A; 4574 C; 4718 G; 4285 T;

Query Match 77.9%; Score 14.8; DB 1; Length 17410;
Best Local Similarity 88.9%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0.

OY 1 AAGGATCAGAAACAATGCC 18
DB 1312 AAGGAGCAGAAAAAATGCC 1295

RESULT 5
U/over
ID Q76820 standard; DNA; 218 BP.
AC Q76820;
DT 23-SEP-1994 (first entry)
DE Human genome fragment.
KW Brain; placenta; bone marrow; genetic analysis; gene mapping;
KW detection; homology; human; adrenal tissue; ds.
OS Homo sapiens.
PN W09401548-A.
PD 20-JAN-1994.
PE 13-JUL-1993; G01467.
PR 13-JUL-1992; GB-014657.
PA (MEDIC-) MEDICAL RES COUNCIL.
PA Gross J, Hadfield KM, Howells D, Kelly M, Shaw D;

```

PI	Sibson DR; Starkey M;
DR	WPL: 94-035056/04.
PT	New nucleic acid fragment encoding gene products - can be used
PT	for genetic analysis and mapping
PS	Claim 1; Page 222; 616pp; English.
CC	Human nucleic acid fragments, isolated from brain, adrenal tissue,
CC	the placenta or bone marrow comprise any of: (A) a sequence
CC	selected from (Q76401-Q77613), (B) an allelic variation of a
CC	sequence as described in (A), or (C) a sequence complementary
CC	to (A) or (B).
SC	Sequence 218 BP; 45 A; 68 C; 55 G; 50 T;
OY	1 AACGATCAGAACAAATGCC 18
DB	94 AAGGATACGATCAATGCC 111
Query Match	77.9%; Score 14.8; DB 1; Length 218;
Best Local Similarity	88.9%; Pred. No. 56;
Matches 16; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
RESULT	6
ID	Q72708/c
AC	Q72708 standard; DNA: 17350 BP.
DC	Q72708:
DE	09-JUN-1995 (first entry)
KM	hop1: human osteogenic protein 1 partial genomic DNA.
KM	hop1: human osteogenic protein; osteoarthritis; orthogenesis;
KM	non-union fracture repair; allograft repair;
KM	cartilage and endochondrial bone formation;
OS	periodontal, dental and craniofacial reconstruction; ss.
OS	Homo sapiens
FH	Key
FH	Location/Qualifiers
FT	exon 3192..3730
FT	/tag- a
FT	/label- EXON-1
FT	10413..10414
FT	/tag- b
FT	/label- GAP-1
FT	/note- "a gap occurs between positions 10413 and 10414
FT	in this sequence"
FT	10696..10891
FT	/tag- c
FT	/label- EXON-2
FT	10960..10961
FT	/tag- d
FT	/label- GAP-2
FT	/note- "a gap occurs between positions 10960 and 10961
FT	in this sequence"
FT	11059..11211
FT	/tag- e
FT	/label- EXON-3
FT	11351..11352
FT	/tag- f
FT	/label- GAP-3
FT	/note- "a gap occurs between positions 11351 and 11352
FT	in this sequence"
FT	11420..11617
FT	/tag- g
FT	/label- EXON-4
FT	11721..11722
FT	/tag- h
FT	/label- GAP-4
FT	/note- "a gap occurs between positions 11721 and 11722
FT	in this sequence"
FT	13354..13436
FT	/tag- i
FT	/label- EXON-5
FT	15044..15160
FT	/tag- j
FT	/label- EXON-6
FT	17245..17410
FT	exon

```
FT      /*tag- k
FT      /label- EXON-7
PN      US5354557-A.
PD      11-OCT-1994.
PF      08-APR-1988; 179406.
PR      08-APR-1988; US-179406.
PR      15-AUG-1988; US-223630.
PR      23-FEB-1989; US-315342.
PR      17-OCT-1989; US-422613.
PR      17-OCT-1989; US-422699.
PR      22-FEB-1990; US-4683913.
PR      20-AUG-1990; US-569920.
PR      07-SEP-1990; US-579865.
PR      18-OCT-1990; US-599543.
PR      18-OCT-1990; US-600024.
PR      04-DEC-1990; US-621849.
PA      04-DEC-1990; US-621988.
PA      22-FEB-1991; US-660162.
PA      20-DEC-1991; US-810560.
PA      28-JUN-1992; US-827052.
PA      21-FEB-1992; US-841646.
PA      18-DEC-1992; US-993387.
PI      Kuberasampath T, Oppermann H, Ozkaynak E, Pang RHL;
PI      Rueger DC;
DR      MPI: 94-324521/40.
PT      Implantable device for inducing osteogenesis - comprises porous
PT      matrix contg. non-glycosylated dimeric, di:sulphide linked
PT      osteogenic protein.
PS      Disclosure: Columns 71-86; 128pp; English.
CC      072708 is the human osteogenic protein 1 (hop1) genomic DNA sequence,
CC      the protein-coding region for hop1 (R51644) is encoded in the seven
CC      exons. Fragments of this protein consisting of residues 335-431,
CC      318-431, 293-431, 300-431, 313-431, 315-431 and 316-431; are
CC      unglycosylated osteogenic polypeptides. Any two of these polypeptides
CC      can be disulphide bonded to form a dimer, which forms an essential
CC      component of an osteogenic protein. This protein is dispersed
CC      in a biodegradable matrix which can be implanted into a mammalian
CC      bone marrow cavity, here it can induce local cartilage, bone and
CC      endochondral bone formation; and it can also accelerate allograft
CC      repair. This implant has the advantage of inducing all stages of
CC      bone formation and of having a higher specific activity than other
CC      known biosynthetic materials. The implant can be used to repair
CC      non-union fractures and cartilage; treat osteoarthritis; and aid
CC      in periodontal, dental or craniofacial reconstruction.
SQ      Sequence 17350 BP; 3813 A; 4569 C; 4688 G; 4280 T;

Query Match      77.9%; Score 14.8; DB 1; Length 17350;
Best Local Similarity 88.9%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGATCAGAACATGCC 18
Dh      1312 AAGGACGAGAAAATGCC 1295

RESULT 7
ID      T11639 standard; DNA; 17415 BP.
AC      T11639;
CT      16-APR-1996 (first entry)
DE      Human osteogenic protein-1 intron non-coding sequence.
KW      Osteogenic protein; OP-1; reporter gene; screening; Identification;
KW      Intron; non-coding sequence; ss.
CN      Homo sapiens.
PN      WO9533831-A1.
PD      14-DEC-1995.
PF      07-JUN-1995; U07349.
PR      07-JUN-1994; US-255250.
PA      (CREA-) CREATIVE BIOMOLECULES INC.
PI      Oppermann H, Ozkaynak E;
DR      MPI: 96-040236/04.
PT      Isolation of compounds to modulate OP-1 expression - by monitoring
```

```
PT      expression changes in a cell transformed to express osteogenic
PT      protein-1 and having additional steroid binding site
PS      Claim 3; Page 43-53; 77pp; English.
CC      The human osteogenic protein-1 (OP-1) non-coding sequence is used in
CC      the construction of an expression vector comprising a reporter gene
CC      which has the non-coding sequence lying contiguous to the reporter
CC      gene and which is able to act on and affect the expression of
CC      the reporter gene when bound to by candidate compounds. The method
CC      is used to identify compounds capable of modulating OP-1 expression.
SQ      Sequence 17415 BP; 3826 A; 4581 C; 4723 G; 4285 T;

Query Match      77.9%; Score 14.8; DB 1; Length 17415;
Best Local Similarity 88.9%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGATCAGAACATGCC 18
Dh      1314 AAGGACGAGAAAATGCC 1297

RESULT 8
ID      T118381/c
ID      T118381;
DE      20-JUN-1996 (first entry)
DE      hop-1 genomic DNA.
KW      Human; osteogenic protein; hop-1; murine; mop-1; TGF-beta superfamily;
KW      transforming growth factor-beta; dimer; antibody; epiloque; hippocampus;
KW      purification; implantable osteogenic device; bone formation; craniofacial;
KW      anomaly; skeletal; dental; endochondral bone formation;
KW      non-union fracture; cartilage repair; osteoarthritis; ss.
OS      Homo sapiens.
FH      Key
FH      Location/Qualifiers
FT      exon
FT      3192..3730
FT      /tag= a
FT      /number= 1
FT      3313..3315
FT      /tag= b
FT      /function= Start codon
FT      10413..10414
FT      /tag= c
FT      /note= "Position of possible nucleotide deletion"
FT      10697..10889
FT      /tag= d
FT      /number= 2
FT      11020..11021
FT      /tag= e
FT      /note= "Position of possible nucleotide deletion"
FT      11063..11210
FT      /tag= f
FT      /number= 3
FT      11351..11352
FT      /tag= g
FT      11421..11618
FT      /note= "Position of possible nucleotide deletion"
FT      /tag= h
FT      /number= 4
FT      11598..11599
FT      /tag= i
FT      /note= "Position of deletion of G compared to CDNA
FT      sequence given in T02597"
FT      11618
FT      /tag= j
FT      /note= "Position of insertion of G compared to CDNA
FT      sequence given in T02597"
FT      11721..11722
FT      /tag= k
FT      13358..13435
FT      /tag= l
FT      /number= 5
FT      15047..15157
FT      /tag= m
FT      exon
```

FT /number- 5
FT 17246. .17410
FT /*tag- n
FT /number- 7
FT US5468845-A.
PN 21-NOV-1995.
PD 08-APR-1988; 179406.
PR 08-APR-1988; US-179406.
PR 15-AUG-1988; US-332630.
PR 23-FEB-1989; US-315342.
PR 17-OCT-1989; US-422699.
PR 22-FEB-1989; US-483913.
PR 20-AUG-1990; US-569920.
PR 07-SEP-1990; US-579865.
PR 18-OCT-1990; US-599543.
PR 21-NOV-1990; US-616374.
PR 04-DEC-1990; US-621849.
PR 22-FEB-1991; US-621988.
PR 20-DEC-1991; US-660162.
PR 28-JAN-1992; US-810560.
PR 21-FEB-1992; US-827052.
PR 01-NOV-1993; US-147023.
PA (STRYKER) STRYKER CORP
PI Kuderzampath T, Oppermann H, Ozkaynak E, Rueger DC;
DR WPI: 96-010159/01.
P-PSDB: R85757.
PT Antibodies with osteogenic protein binding specificity - used in
P1 purification of osteogenic proteins, and as antigenic proteins
PS Disclosure: Column 71-86; 129pp; English.
CC This sequence represents the genomic sequence which encodes the human
CC osteogenic protein, hOP-1. hOP-1 has homology with proteins in the
CC TGF-beta superfamily. Fragments of the hOP-1 protein can be used in
CC the production of dimeric peptides which may be used in the generation
CC of antibodies with binding specificities for osteogenic proteins. The
CC antibodies are capable of binding specifically to an epitope of the
CC osteogenic protein and may be used in purification protocols. Osteogenic
CC proteins, such as hOP-1, may be used in an implantable osteogenic device
CC which allows predictable bone formation to correct acquired and
CC congenital craniofacial and other skeletal or dental anomalies. They may
CC be used to induce local endochondral bone formation in non-union
CC fractures and in other clinical applications including dental and
CC periodontal applications where bone formation is required. Other
CC potential applications include cartilage repair, e.g. in the treatment
CC of osteoarthritis.
SQ Sequence 17410 BP; 3831 A; 4576 C; 4716 G; 4287 T;
Query Match 77.9%; Score 14.8; DB 1; Length 17410;
Best Local Similarity 88.9%; Pred. No. 94;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AAGGATCAGACATGCC 18
DB 1312 AAGGAGCAGAAAAATGCC 1295
RESULT 9
ID T58758 standard; DNA: 3931 BP.
AC T58758
DT 20-MAR-1997 (first entry)
DE Murine long form CSF-1 coding sequence.
KW Long form; short form; human colony stimulating factor-1; HuCSF-1;
KW truncated protein; immunosuppression; chemotherapy;
KW bone marrow transplantation; AIDS; ss.
OS Mus musculus.
FH Key
FH Location/Qualifiers
FT cds
FT /*tag- a
FT /*Product= Full_length_long_form_mucsf-1
FT /*signal_peptide 160..255

FT /*tag- b
FT mat_peptide 256..1815
FT /*tag- c
FT US5573930-A.
PN 12-NOV-1996.
PD 05-FEB-1985; 698359.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-823067.
PR 16-APR-1987; US-039657.
PR 16-APR-1987; US-039654.
PR 27-NOV-1991; US-799411.
PR 27-NOV-1991; US-799039.
PR 28-DEC-1992; US-999288.
PA (CETUS) CETUS ONCOLOGY CORP.
PI Coyne MY, Halenbeck RF, Kawasaki ES, Kolts RE, Ladner MB;
PI Martin GA, Noble JA;
DR WPI: 96-517883/51.
P-PSDB: W10070.
PT DNA encoding human colony-stimulating factor-1 N-terminal deletion
PT mutants - opt. with C-terminal truncation(s), useful for treating
PT immunosuppression caused by e.g. chemotherapy or bone marrow
PT transplants
PS Disclosure: Column 61-68; 45pp; English.
CC The sequences given in T58758-59 encode the long and short forms of
CC murine colony stimulating factor (mucsf)-1. These sequences are closely
CC related to human CSF-1 sequences which were used in the design of the
CC truncated proteins of the invention (see also W10072-86). These
CC truncated proteins have 3 amino acids deleted at the N-terminal end,
CC and a variable number deleted from the C-terminal end. The truncated
CC proteins have mol. wts. much closer to those found in naturally
CC occurring CSF-1 dimers, and it is thought that natural CSF-1 may be
CC C-terminally truncated. The novel proteins may be useful for overcoming
CC immunosuppression induced e.g. by chemotherapy, bone marrow
CC transplantation or diseases, e.g. AIDS.
SQ Sequence 3931 BP; 973 A; 1106 C; 1004 G; 848 T;
Query Match 77.9%; Score 14.8; DB 1; Length 3931;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 AAGATCAGAACATGCCT 19
DB 542 AAGAGCAGAACAGGCCT 559
RESULT 10
ID T58759 standard; DNA: 1987 BP.
AC T58759
DT 20-MAR-1997 (first entry)
DE Murine short form CSF-1 coding sequence.
KW Long form; short form; human colony stimulating factor-1; HuCSF-1;
KW truncated protein; immunosuppression; chemotherapy;
KW bone marrow transplantation; AIDS; ss.
OS Mus musculus.
FH Key
FH Location/Qualifiers
FT cds
FT /*tag- a
FT /*tag- b
FT /*tag- c
FT mat_peptide 166..1725
FT /*tag- b
FT /*tag- c
FT US5573930-A.
PN 12-NOV-1996.
PD 05-FEB-1985; 698359.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-823067.
PR 16-APR-1987; US-039657.
PR 16-APR-1987; US-039654.
PR 27-NOV-1991; US-799411.
PR 27-NOV-1991; US-799039.
PR 28-DEC-1992; US-999288.
PA (CETUS) CETUS ONCOLOGY CORP.
PI Coyne MY, Halenbeck RF, Kawasaki ES, Kolts RE, Ladner MB;
PI Martin GA, Noble JA;
DR WPI: 96-517883/51.
P-PSDB: W10070.
PT DNA encoding human colony-stimulating factor-1 N-terminal deletion
PT mutants - opt. with C-terminal truncation(s), useful for treating
PT immunosuppression caused by e.g. chemotherapy or bone marrow
PT transplants
PS Disclosure: Column 61-68; 45pp; English.
CC The sequences given in T58758-59 encode the long and short forms of
CC murine colony stimulating factor (mucsf)-1. These sequences are closely
CC related to human CSF-1 sequences which were used in the design of the
CC truncated proteins of the invention (see also W10072-86). These
CC truncated proteins have 3 amino acids deleted at the N-terminal end,
CC and a variable number deleted from the C-terminal end. The truncated
CC proteins have mol. wts. much closer to those found in naturally
CC occurring CSF-1 dimers, and it is thought that natural CSF-1 may be
CC C-terminally truncated. The novel proteins may be useful for overcoming
CC immunosuppression induced e.g. by chemotherapy, bone marrow
CC transplantation or diseases, e.g. AIDS.
SQ Sequence 3931 BP; 973 A; 1106 C; 1004 G; 848 T;

PR 30-APR-1985; US-728834.
PR 14-JUN-1983; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-039657.
PR 16-APR-1987; US-039654.
PR 27-NOV-1991; US-799411.
PR 27-NOV-1991; US-799039.
PR 28-DEC-1992; US-999298.
PA (CETU) CERUS ONCOLOGY CORP.
PI Coyne MY, Halenbeck RF, Kawasaki ES, Kolts KE, Ladner MB;
PI Martin GA, Noble JA;
DR WPI; 96-51783/51.
P-PSDB; W10071.
PT DNA encoding human colony-stimulating factor-1 N-terminal deletion
PT mutants - opt. with C-terminal truncation(s), useful for treating
PT immunosuppression caused by e.g. chemotherapy or bone marrow
PT transplants
PS Disclosures: Column 71-76; 45pp; English.
CC The sequences given in T58758-59 encode the long and short forms of
CC murine colony stimulating factor (mCSF)-1. These sequences are closely
CC related to human CSF-1 sequences which were used in the design of the
CC truncated proteins of the invention (see also W10072-86). These
CC truncated proteins have 3 amino acids deleted at the N-terminal end,
CC and a variable number deleted from the C-terminal end. The truncated
CC proteins have mol. wts. much closer to those found in naturally
CC occurring CSF-1 dimers, and it is thought that natural CSF-1 may be
CC C-terminally truncated. The novel proteins may be useful for overcoming
CC immunosuppression induced e.g. by chemotherapy, bone marrow
CC transplantation or diseases, e.g. AIDS.
SQ Sequence 1987 BP; 496 A; 569 C; 511 G; 411 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1987;
Best Local Similarity 88.9%; Pred. No. 73;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGATCAGAACATGCCT 19
11111111111111111111

DB 452 AGGAGCAGAACAGGCCT 469

RESULT 11
T47073
ID T47073 standard; RNA; 8543 BP.
AC T47073;
DT 15-APR-1997 (first entry)
DE Maize dwarf mosaic virus genomic RNA.
KW MDV-B; viral resistance; disease resistance; transgenic plant;
KW monoco; P3 proteinase; N1a proteinase; N1b replicase;
KW RNA-dependent RNA polymerase; coat protein; Zea mays; sorghum;
KW sugarcane; Saccharum officinale; ss.
OS Maize dwarf mosaic virus strain B.
FH Key location/Qualifiers
FT cds 3..8291
FT /tag- a
FT /product- polypeptide encoded by MDV-B genome
FT 3'utr 8292..8530
FT /tag- b
FT 3..1133
FT /tag- c
FT /product- 3-prime sequence of helper component
FT proteinase (HC-Pro)
FT 1134..2375
FT /tag- d
FT /product- P3 proteinase
FT 2376..4292
FT /tag- e
FT /product- cylindrical inclusion protein
FT 4293..4451
FT /tag- f

FT /product- K2 (6 kDa protein)
FT 4452..5744
FT /tag- g
FT /product- N1a proteinase
FT 5745..7307
FT /tag- h
FT /product- N1b replicase
FT 7308..8291
FT /tag- i
FT /product- coat protein

PN W09702352-A1.
PD 23-JAN-1997.
PE 20-JUN-1996; E02673.
PR 30-JUN-1995; US-496944.
PA (CIBA) CIBA GEIGY AG.
PI Dietz JM, Law MD;
DR WPI; 97-108965/10.
P-PSDB; W10344.
PT Chimeric gene for imparting viral resistance to plants - contains
PT sequence modified to express non-translatable mRNA, or non-coat
PT viral protein
PS Disclosures: Page 31-44; 64pp; English.
CC The sequence of the polycistronic messenger RNA of maize dwarf
CC mosaic virus strain B (MDV-B) is given in T47073 and the encoded
CC MDV-B polypeptide in W10344. New chimeric genes (see also
CC T47074) comprise a monocytledonous plant promoter linked to a
CC modified nucleic acid sequence derived from the MDV-B genome. The
CC modification is such that mRNA is translated from a truncated protein
CC (pref. smaller than 200 amino acids), no translation of mRNA occurs
CC or the transcribed mRNA lacks the translation initiation codon or
CC includes a premature stop codon. Expression of the chimeric gene
CC inhibits infection of plants (pref. sorghum, sugarcane, esp. maize)
CC by MDV. The transgenic plants display an inheritable resistance
CC trait.

SQ Sequence 8543 BP; 2913 A; 1637 C; 1820 G; 2160 U;

Query Match 77.9%; Score 14.8; DB 1; Length 8543;
Best Local Similarity 72.2%; Pred. No. 87;
Matches 13; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 AGGATCAGAACATGCCT 19
11111111111111111111

DB 4526 AGGATCGAAGAAUGCCU 4543

RESULT 12
T47074
ID T47074 standard; RNA; 1296 BP.
AC T47074;
DT 15-APR-1997 (first entry)
DE Maize dwarf mosaic virus N1a protease RNA.
KW MDV-B; viral resistance; disease resistance; transgenic plant;
KW monoco; P3 proteinase; N1a proteinase; N1b replicase;
KW RNA-dependent RNA polymerase; coat protein; Zea mays; sorghum;
KW sugarcane; Saccharum officinale; maize dwarf mosaic virus; ss.
OS Synthetic.
FH Key location/Qualifiers
FT mutation 18
FT /tag- a
FT /note- "Substn. of T for A at position 18 creates
FT a premature stop codon"

PN W09702352-A1.
PD 23-JAN-1997.
PE 20-JUN-1996; E02673.
PR 30-JUN-1995; US-496944.
PA (CIBA) CIBA GEIGY AG.
PI Dietz JM, Law MD;
DR WPI; 97-108965/10.
PT Chimeric gene for imparting viral resistance to plants - contains
PT sequence modified to express non-translatable mRNA, or non-coat
PT viral protein
PS Claim 14; Page 38-40; 64pp; English.

CC N1a protease RNA (T47074) isolated from maize dwarf mosaic virus
(MDMV) (see also T47073) was amplified by PCR and modified to
include an ATG initiation codon and premature stop codon. The
CC construct was inserted into pCIB4421, contg. the PEP promoter and
CC CMV 35S terminator to give pCIB5018. The plasmid was used to
CC transform maize line CG00526 embryos. Plants that developed from
CC the embryos were screened for the transgene by PCR (see also
CC T47075-80). Seeds from selected lines were grown to plants and
CC these exposed to MDMV. Some plants were resistant, i.e. they
CC contained no detectable N1a transcripts and no virus.
SQ Sequence 1296 BP; 450 A; 252 C; 276 G; 2 T; 316 U;

Query Match 77.9%; Score 14.8; DB 1; Length 1296;
Best Local Similarity 72.2%; Pred. No. 69;
Matches 13; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 2 AGGATCAGACATGCT 19
|||:| |||:|:|:
Db 78 AGGATCAGACATGCT 95

RESULT 13
ID T80599 standard; cDNA; 3931 BP.
AC T80599: (first entry)
DT 07-NOV-1997
DE Murine 4 kb colony stimulating factor-1 encoding cDNA.
KW CSF-1; macrophage colony stimulating factor; M-CSF; immunostimulant;
KW immunosuppression; chemotherapy; transplant; AIDS; neoplasm;
KW acquired immune deficiency syndrome; antitumour; lymphokines;
KW mouse; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT signal_peptide 160..255
FT /*tag- a
FT mat_peptide 256..1816
FT /*tag- b
FT /product- CSF-1
PN US5643563-A.
PD 01-JUL-1997.
PE 05-FEB-1985; 698359.
PF 28-DEC-1992; US-999280.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-039654.
PR 16-APR-1987; US-039657.
PR 13-OCT-1987; US-105261.
PR 27-NOV-1991; US-799039.
PR 27-NOV-1991; US-799411.
PR 21-APR-1995; US-426036.
PA (CHIR) CHIRON CORP
PI Coyne MY, Halsebeck RF, Kawasaki ES, Kohls KE, Ladner MB;
PI Martin GA, Noble JA;
PI WPI: 97-350183/32.
DR P-PSDB: W22615.
PT Carboxy-truncated colony-stimulating factor 1 polypeptide(s) - as
PS Immunostimulants for treating immunosuppression
PS Disclosure: Column 63-70; 45pp; English.
CC The present sequence encodes a novel murine 4 kb colony stimulating
CC factor (CSF-1). Human CSF-1 (also known as macrophage colony stimulating
CC factor, M-CSF) proteins are useful as immunostimulants for treating
CC immunosuppressed patients, e.g. chemotherapy patients, transplant
CC recipients or AIDS (acquired immune deficiency syndrome) patients,
CC or for treating neoplasms and infections. They may be administered
CC together with antitumour agents or lymphokines.
SQ Sequence 3931 BP; 973 A; 1106 C; 1004 G; 848 T;

Query Match 77.9%; Score 14.8; DB 1; Length 3931;
Best Local Similarity 88.9%; Pred. No. 79;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 AGGATCAGACATGCT 19
||||| ||||| |||||
Db 542 AGGATCAGACATGCT 559

RESULT 14
ID T80600 standard; cDNA; 1987 BP.
AC T80600: (first entry)
DT 07-NOV-1997
DE Murine 2 kb colony stimulating factor-1 encoding cDNA.
KW CSF-1; macrophage colony stimulating factor; M-CSF; immunostimulant;
KW immunosuppression; chemotherapy; transplant; AIDS; neoplasm;
KW acquired immune deficiency syndrome; antitumour; lymphokines;
KW mouse; ss.
OS Mus musculus.
FH Key Location/Qualifiers
FT signal_peptide 70..165
FT /*tag- a
FT mat_peptide 166..1726
FT /*tag- b
FT /product- CSF-1
PN US5643563-A.
PD 01-JUL-1997.
PE 05-FEB-1985; 698359.
PF 28-DEC-1992; US-999280.
PR 05-FEB-1985; US-698359.
PR 30-APR-1985; US-728834.
PR 14-JUN-1985; US-744924.
PR 18-JUL-1985; US-756814.
PR 21-JAN-1986; US-821068.
PR 20-JUN-1986; US-876819.
PR 24-OCT-1986; US-923067.
PR 16-APR-1987; US-039654.
PR 16-APR-1987; US-039657.
PR 13-OCT-1987; US-105261.
PR 27-NOV-1991; US-799039.
PR 27-NOV-1991; US-799411.
PR 21-APR-1995; US-426036.
PA (CHIR) CHIRON CORP
PI Coyne MY, Halsebeck RF, Kawasaki ES, Kohls KE, Ladner MB;
PI Martin GA, Noble JA;
PI WPI: 97-350183/32.
DR P-PSDB: W22616.
PT Carboxy-truncated colony-stimulating factor 1 polypeptide(s) - as
PS Immunostimulants for treating immunosuppression
PS Disclosure: Column 71-76; 45pp; English.
CC The present sequence encodes a novel murine 2 kb colony stimulating
CC factor (CSF-1). Human CSF-1 (also known as macrophage colony stimulating
CC factor, M-CSF) proteins are useful as immunostimulants for treating
CC immunosuppressed patients, e.g. chemotherapy patients, transplant
CC recipients or AIDS (acquired immune deficiency syndrome) patients,
CC or for treating neoplasms and infections. They may be administered
CC together with antitumour agents or lymphokines.
SQ Sequence 1987 BP; 496 A; 569 C; 511 G; 411 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1987;
Best Local Similarity 88.9%; Pred. No. 73;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
OY 2 AGGATCAGACATGCT 19
||||| ||||| |||||
Db 452 AGGATCAGACATGCT 469

RESULT 15
ID T94752:

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds

(without alignments)
22.116 Million cell updates/sec

Title: US-09-325-095-24

Perfect score: 19

Sequence: 1 AAGGATCAGACAAATGCCT 19

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:
1: em_est1.*
2: em_est2.*
3: em_est3.*
4: em_est4.*
5: em_est5.*
6: em_est6.*
7: em_est7.*
8: em_est8.*
9: em_est9.*
10: em_est10.*
11: em_est11.*
12: em_est12.*
13: em_est13.*
14: em_est14.*
15: em_est15.*
16: em_est16.*
17: em_est17.*
18: em_est18.*
19: em_est19.*
20: em_est20.*
21: em_est21.*
22: em_est22.*
23: em_est23.*
24: em_est24.*
25: em_est25.*
26: em_est26.*
27: em_est27.*
28: em_est28.*
29: em_est29.*
30: em_est30.*
31: em_est31.*
32: em_est32.*
33: em_est33.*
34: em_est34.*
35: em_est35.*
36: em_est36.*
37: em_est37.*
38: em_est38.*
39: em_est39.*
40: em_est40.*
41: em_est41.*
42: em_est42.*
43: em_est43.*
44: em_est44.*
45: em_est45.*
46: em_est46.*
47: em_est47.*
48: em_est48.*
49: em_est49.*
50: em_est50.*
51: em_est51.*
52: em_est52.*
53: em_est53.*

54: em_est22.*
55: em_est23.*
56: em_est24.*
57: em_est25.*
58: em_est26.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.4	91.6	417	29	AA190409	AA190409 zp89c10.r
2	16.4	86.3	479	37	AA731088	AA731088 nv68a08.s
3	16.4	86.3	485	38	AA767230	AA767230 n280h01.s
4	16.4	86.3	918	38	AA767233	AA767233 n280h05.s
5	16.4	86.3	561	41	AI017576	AI017576 ou23f12.x
6	16.4	84.2	634	29	AA151764	AA151764 zo3f06.s
7	16.4	84.2	522	35	AA581298	AA581298 nd43e06.s
8	16.4	84.2	509	47	AI475209	AI475209 t180b03.x
9	16.4	84.2	468	47	AI499908	AI499908 t003f05.x
10	16.4	84.2	414	49	AI640828	AI640828 t269a10.x
11	15.8	83.2	651	29	AA142034	AA142034 CR00149.3
12	15.8	83.2	406	29	AA146711	AA146711 zo35e10.r
13	15.8	83.2	235	31	AA288447	AA288447 vb37a12.r
14	15.8	83.2	403	35	AA593545	AA593545 nm28c07.s
15	15.8	83.2	578	37	AA672725	AA672725 vp07d01.r
16	15.8	83.2	488	38	AA759578	AA759578 vv55c04.r
17	15.8	83.2	260	41	AI048755	AI048755 ud32a10.r
18	15.8	83.2	638	47	AI444697	AI444697 486015c11
19	15.8	83.2	276	47	AI462631	AI462631 vb37a12.x
20	15.8	83.2	641	47	AI467712	AI467712 486015c11
21	15.8	83.2	468	48	AI596476	AI596476 v174h10.x
22	15.8	83.2	491	54	HSN010092	AI045242 homo sapi
23	15.4	81.1	504	38	AA760081	AA760081 vv69c06.r
24	15.4	81.1	500	39	AA898726	AA898726 NCM6E107
25	15.4	81.1	509	42	AI142259	AI142259 qg62d07.r
26	15.4	81.1	294	29	AA175134	AA175134 ms82g07.r
27	15.4	78.9	709	41	AI054610	AI054610 comu0011
28	15.4	78.9	372	46	AI448592	AI448592 mc50b09.x
29	15.4	78.9	370	46	AI450814	AI450814 ms82g07.x
30	15.4	78.9	370	49	AI639779	AI639779 ms82g07.y
31	14.8	77.9	218	20	220318	220318 HSAABXHU P
32	14.8	77.9	286	20	221098	221098 HSAABMSX T
33	14.8	77.9	320	21	F12667	F12667 HSC3CG061.D
34	14.8	77.9	429	21	T74316	T74316 YG84d06.r1
35	14.8	77.9	358	21	T78190	T78190 yd79d12.r1
36	14.8	77.9	479	24	H75125	H75125 250 P1RFG1
37	14.8	77.9	304	26	W20869	W20869 mb96c04.r1
38	14.8	77.9	488	26	W45995	W45995 mc81g05.r1
39	14.8	77.9	486	27	AA021817	AA021817 mb84a01.r
40	14.8	77.9	481	29	AA187027	AA187027 zp72g01.r
41	14.8	77.9	380	30	AA240477	AA240477 mv34b01.r
42	14.8	77.9	340	30	AA247156	AA247156 chp035.s
43	14.8	77.9	482	30	AA259669	AA259669 va54d09.r
44	14.8	77.9	440	30	AA260212	AA260212 va88f07.r
45	14.8	77.9	652	53	HSN007439	AI042589 Homo sapi

ALIGNMENTS

RESULT 1
AA190409
LOCUS
DEFINITION
IMAGE:627378 5' similar to SW:P11A.BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.
AA190409
ACCESSION
AA190409
417 bp
mRNA
EST
15-JAN-1997
zp89c10.r1 Striatagene Hela cell s3 937216 Homo sapiens cDNA clone
IMAGE:627378 5' similar to SW:P11A.BOVIN P32871
PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;
mRNA sequence.

NID 61779240
 VERSION AA190409.1 GI:1779240
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 417)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapelli, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Maritz, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rikkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478
 COMMENT On May 18, 1995 this sequence version replaced gi:811202.

TITLE
 JOURNAL
 MEDLINE
 COMMENT

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewartson.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 328.

FEATURES
 source
 1..417
 /organism="Homo sapiens"
 /db_xref="GDB:5046076"
 /db_xref="taxon:9606"
 /clone="IMAGE:627378"
 /clone_1ib="NCI-CGAP.GCB1"
 /clone_1ib="Stratagene Hela cell 83 937216"
 /sex="female"
 /dev_stage="Hela S3 cell line"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Hela S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR vector. -5' adaptor sequence: 5' GAATTCGGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 116 a 98 c 93 g 99 t 11 others
 ORIGIN

Query Match 91.6%; Score 17.4; DB 29; Length 417;
 Best Local Similarity 94.7%; Pred. No. 38;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AAGATCAGACAAATGCT 19
 ||| ||||| ||||| |||||
 DB 172 AAGATCAGACAAATGCT 190

RESULT 2
 AA731088 479 bp mRNA EST 07-FEB-1998
 LOCUS n068001.s1 NCI-CGAP.GCB1 Homo sapiens cDNA clone IMAGE:1251734 3'
 DEFINITION similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170 KDA SUBUNIT. ;
 mRNA sequence.
 ACCESSION AA731088
 NID 92752292
 VERSION AA731088.1 GI:2752292
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 479)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 TITLE Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1395051.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/bdrrp/image/image.html

FEATURES
 source
 1..479
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1251734"
 /clone_1ib="NCI-CGAP.GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI). Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - Oligo(dT) primer.
 15'-TGTACCAATCTGAAGTGGAGCGCCGCTCATTTTCTTTTCTTTT-3'
 3'-Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified p773 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 85 a 122 c 108 g 164 t
 ORIGIN

Query Match 86.3%; Score 16.4; DB 37; Length 479;
 Best Local Similarity 94.4%; Pred. No. 1.2e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGATCAGACAAATGCT 19
 ||| ||||| ||||| |||||
 DB 436 AGGATCAGACAAATGCT 419

RESULT 3
 AA767230 485 bp mRNA EST 08-FEB-1998
 LOCUS n280h01.s1 NCI-CGAP.GCB1 Homo sapiens cDNA clone IMAGE:1301809 3'
 DEFINITION similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170 KDA SUBUNIT. ;
 mRNA sequence.
 ACCESSION AA767230
 NID 92818245
 VERSION AA767230.1 GI:2818245
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150988.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert length: 1003 Std Error: 0.00
Seq primer: -40ml3 fwd: ET from Amersham.

FEATURES
source Location/Qualifiers

1..485
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="21g"
/clone="IMAGE:1301809"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGTGGAGGCGCCCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 87 a 125 c 107 g 166 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 38; Length 485;
Best Local Similarity 94.4%; Pred. NO.1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGATCAGAACATGCCT 19
||||| |||||||||
DB 435 AGGATAAGAACATGCCT 418

RESULT 4
AA767233 918 bp mRNA EST 08-FEB-1998
LOCUS n280h05.s1 NCI CGAP GCB1 Homo sapiens cDNA clone IMAGE:1301817 3'
DEFINITION similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170 KDA SUBUNIT. ;,
mRNA sequence.
ACCESSION AA767233
NID 92818248
VERSION AA767233.1 GI:2818248
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 918)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2150992.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert length: 1580 Std Error: 0.00
Seq primer: -40ml3 fwd: ET from Amersham
High quality sequence stop: 484.

FEATURES
source Location/Qualifiers

1..918
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11; 21g"
/clone="IMAGE:1301817"
/clone_lib="NCI-CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"

/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAGTGGAGGCGCCCTCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 194 a 220 c 230 g 274 t
ORIGIN

Query Match 86.3%; Score 16.4; DB 38; Length 918;
Best Local Similarity 94.4%; Pred. NO.1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGGATCAGAACATGCCT 19
||||| |||||||||
DB 435 AGGATAAGAACATGCCT 418

RESULT 5
A1017576 561 bp mRNA EST 27-AUG-1998
LOCUS 0233f12.x1 Soares NFI_T GBC S1 Homo sapiens cDNA clone
DEFINITION IMAGE:1627151 3' similar to TR:Q92923 Q92923 SWI/SNF COMPLEX 170
KDA SUBUNIT. ;, mRNA sequence.
ACCESSION A1017576
NID 93231912
VERSION A1017576.1 GI:3231912
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 561)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT On Jan 14, 1998 this sequence version replaced gl:1798155.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 988 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 477.
Location/Qualifiers

FEATURES

source

```
1. 561
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pRTT3D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NBHL19W, testis NHT, and B-cell
NCI-CGAP, GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
726408-728711, and 729096-733399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
141 c 127 g 182 t
```

BASE COUNT
ORIGIN

Query Match 86.3%; Score 16.4; DB 41; Length 561;
Best Local Similarity 94.4%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 AGCATCGAACAATGCTT 19
||||| |||||||||
DB 428 AGGATAGAACAAATGCTT 411

RESULT 6
AA151764/c 634 bp mRNA EST 06-AUG-1997
LOCUS
DEFINITION
z30306.s1 StrataGene colan (#937204) Homo sapiens cDNA clone
IMAGE:588419 3' similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL
PROTEIN L3. ; mRNA sequence.
ACCESSION
AA151764
NID
91720478
VERSION
AA151764.1 GI:1720478
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 634)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chapel, J.B.,
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rikkin, L.,
Rohlfing, T., Schellberg, K., Soares, M.B., Tan, E., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 688 Std Error: 0.00
Seq primer: -40m13 fwd. from Amersham
High quality sequence stop: 391.
Location/Qualifiers

FEATURES

source

```
1. 634
/organism="Homo sapiens"
/db_xref="GDB:4620721"
/db_xref="taxon:9606"
/clone_lib="IMAGE:588419"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (Xenopus resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:  
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. T-84 colonic epithelial cell line. Average  
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor  
sequence: 5' GAATTCGGACGACG 3' -3' adaptor sequence: 5'  
CTCGAGTTTCTTTTCTTTTCTTTT 3'"
137 a 169 c 156 g 163 t 9 others
```

BASE COUNT
ORIGIN

Query Match 84.2%; Score 16; DB 29; Length 634;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 GATCAGACAAATGCTT 19
||||| |||||||||
DB 371 GATCAGACAAATGCTT 356

RESULT 7
AA581298 522 bp mRNA EST 05-JAN-1998
LOCUS
DEFINITION
nd43e06.s1 NCI-CGAP_R1 Homo sapiens cDNA clone IMAGE:803074 3'
similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL PROTEIN L3. ; mRNA
sequence.
ACCESSION
AA581298
NID
92359070
VERSION
AA581298.1 GI:2359070
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 522)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 9, 1995 this sequence version replaced gl:804204.

COMMENT
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Lee Helman, Ph.D., Michael Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: Life Technologies Inc., David Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.llnl.gov/bbrp/image/image.html

Seq primer: -4m13 fwd. ET from Amersham
High quality sequence stop: 321.
Location/Qualifiers
1. 522
/organism="Homo sapiens"

FEATURES

source

/db_xref="taxon:9606"
/clone_image:803074"
/clone_id="NCI_CGAP_A1"
/tissue_type="bulk alveolar tumor"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT2; Site_1: SalI; Site_2: NotI;
Cloned unidirectionally. Primer: Oligo dt."
BASE COUNT 114 a 138 c 126 g 144 t
ORIGIN

Query Match 84.2%; Score 16; DB 35; Length 522;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GATCAGAACATGCCT 19
|||||
Db 379 GATCAGAACATGCCT 364

RESULT 8
A1475209/c 509 bp mRNA EST 14-APR-1999
LOCUS t180b03.x1 NCI CGAP Col4 Homo sapiens cDNA clone IMAGE:2153357 3'
DEFINITION similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL PROTEIN L3. ; mRNA
sequence.
ACCESSION A1475209
NID 94328254
VERSION A1475209.1 GI:4328254
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 509)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246645.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 1472 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 316.
Location/Qualifiers
1. 509
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:2153357"
/clone_id="NCI_CGAP_Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT 111 a 135 c 119 g 144 t
ORIGIN

Query Match

84.2%; Score 16; DB 47; Length 509;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GATCAGAACATGCCT 19
|||||
Db 383 GATCAGAACATGCCT 368

RESULT 9
A1499908/c 468 bp mRNA EST 12-MAY-1999
LOCUS t003f05.x1 NCI CGAP ut2 Homo sapiens cDNA clone IMAGE:2177985 3'
DEFINITION similar to SW:RL3_HUMAN P39023 60S RIBOSOMAL PROTEIN L3. ; mRNA
sequence.
ACCESSION A1499908
NID 94391890
VERSION A1499908.1 GI:4391890
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 468)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Mar 20, 1998 this sequence version replaced gi:2979745.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html
Insert Length: 1395 Std Error: 0.00
Seq primer: -40UP from GIBCO
High quality sequence stop: 338
POLYA-No.
Location/Qualifiers
1. 468
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image:2177985"
/clone_id="NCI_CGAP_Ut2"
/tissue_type="moderately-differentiated endometrial
adenocarcinoma, 3 pooled tumors"
/lab_host="DH10B"
/note="Organ: uterus; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.85 kb. Life Technologies catalog #:
11539-012"

BASE COUNT 104 a 120 c 99 g 145 t
ORIGIN

Query Match 84.2%; Score 16; DB 47; Length 468;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 GATCAGAACATGCCT 19
|||||
Db 389 GATCAGAACATGCCT 374

RESULT 10
A1640828/c

Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -28m13 rev2 from Amersham
 High quality sequence stop: 255.
 Location/Qualifiers

FEATURES

Source

1..406
 /organism="Homo sapiens"
 /db_xref="GDB:4626646"
 /db_xref="taxon:9606"
 /clone="IMAGE:588906"
 /clone_lib="Stratagene colon (#937204)"
 /lab_host="SOLR cells (Kanamycin resistant)"
 /note="Organ: colon; Vector: pBluescript SK-; Site: 1:
 EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. T-84 colonic epithelial cell line. Average
 insert size: 1.0 kb; uni-ZAP XR Vector: -5' adaptor
 sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 122 a 83 c 102 g 98 t 1 others

ORIGIN

Query Match 83.2%; Score 15.8; DB 29; Length 406;
 Best Local Similarity 89.5%; Pred. No. 2.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGATCAGACATGCCT 19
 ||||| ||||| ||||| |||||
 Db 318 AAGGACGAGAAAAATGCCT 336

RESULT 13
 AA288447 235 bp mRNA EST 14-APR-1997
 LOCUS VB37A12.1 Soares mouse lymph node NBM1N Mus musculus cDNA clone
 DEFINITION IMAGE:751102 5', mRNA sequence.
 ACCESSION AA288447
 NID 91937377
 VERSION AA288447.1 GI:1937377
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 235)
 AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theisling, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.
 TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On May 17, 1993 this sequence version replaced gi:426136.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 MGI:460086
 putative full length read
 vector to vector length is 239
 Seq primer: -28m13 rev2 ET from Amersham.

FEATURES

Source

1..235
 /organism="Mus musculus"

/strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:751102"
 /clone_lib="Soares mouse lymph node NBM1N"
 /sex="male"
 /tissue="type="lymph node"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Organ: lymph node; Vector: pRT73D-Pac (Pharmacia)
 with a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 [5',
 TGTACCAATGTGAAGTGGAGCGCCGCACTTTTCTTTTCTTTTCTTTTCTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pRT73 vector. RNA
 provided by Dr. Bertard Jordan. Library constructed and
 normalized by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 85 a 48 c 49 g 53 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 31; Length 235;
 Best Local Similarity 89.5%; Pred. No. 2.4e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 AAGGATCAGACATGCCT 19
 ||||| ||||| ||||| |||||
 Db 151 AAGGATCGCAGACATGCCT 169

RESULT 14
 AA593545 403 bp mRNA EST 25-SEP-1997
 LOCUS nm28c07.s1 NCI CGAP Gas1 Homo sapiens cDNA clone IMAGE:1085196 3'
 DEFINITION similar to contains element MCR8 repetitive element; mRNA
 sequence.
 ACCESSION AA593545
 NID 92409307
 VERSION AA593545.1 GI:2409307
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 403)
 AUTHORS NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 9, 1995 this sequence version replaced gi:802451.

Contact: Robert Strussberg, Ph.D.
 Tel: (301) 486-1550
 Email: Robert.Strussberg@nih.gov
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www.bio.lnl.gov/bhrp/image/image.html

Insert length: 956 Std Error: 0.00
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 380.
 Location/Qualifiers

FEATURES

Source

1..403
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="12p"
 /clone="IMAGE:1085196"

```

/clone_lib="NCI_GAP_Gast"
/issue_type="gastric tumor"
/lab_host="SOLR (kanamycin resistant)"
/notes="Organ: stomach; Vector: Bluescript SK-; Site: 1;
EcoRI; Site: 2; XhoI; Cloned unidirectionally. Primer:
oligo dT. Pooled gastric tumors. 5' adaptor sequence: 5'
GAATTCGCGCAG 3' 3' adaptor sequence: 5'
CTTCAGTTTTTTTTTTTTTTT 3' Average insert size: 1.0 kb."
BASE COUNT      112 a      92 c      95 g      104 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 35; Length 403;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGATCAGAACATGCT 19
      ||| | ||||| |||||
DB      12 AAGCACCAGAACATGCT 30

RESULT 15      AA672725      578 bp      mRNA      EST      25-NOV-1997
LOCUS      AA672725/c      VP07601.f1 Soares mouse mammary gland NBMNG Mus musculus cDNA clone
DEFINITION      IMAGE:1067905 5', mRNA sequence.
ACCESSION      AA672725
NID      92644942
VERSION      AA672725.1 GI:2644942
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 578)
Maire, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Schellberg, K., Sepcoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterson, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 19, 1997 this sequence version replaced gi:1520426.

TITLE      The WashU-HMI Mouse EST Project
JOURNAL
COMMENT
Contact: Maria M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:590265
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 498.
Location/Qualifiers
1. 578
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_image="IMAGE:1067905"
/clone_lib="Soares mouse mammary gland NBMNG"
/sex="male"
/issue_type="mammary gland"
/dev stage="4 weeks"
/lab_host="DH10B"
/notes="Organ: mammary gland; Vector: pTT73D-Pac
(Pharmacia) with a modified polylinker; Site: 1;
Site: 2; Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer 15'
TGTACCATCTGAGTGGAGCGCGGCGAATGTTTTTTTTTTTTTTTTTTT
T 3'; double-stranded cDNA was ligated to Eco RI

```

```

adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pTT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT      142 a      121 c      137 g      178 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 37; Length 578;
Best Local Similarity 89.5%; Pred. No. 2.4e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 AAGGATCAGAACATGCT 19
      ||| | ||||| |||||
DB      148 AAGGTCAGAACATGACT 130

Search completed: September 13, 1999, 15:45:56
Job time: 4044 sec

```


CDS

```

/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AAB62534.1"
/db_xref="PID:92245506"
/db_xref="GI:163520"
/translation="MPPRSSGELMGILHMPRIIVCELLPNGMIVTECLREATLT
IKHEFKARKPYLIQIODESSYTFVYTOGEAREEFDETERRCDLRLOPFLKVI
EVRGNERELILREIGFAIGMTCFEDVAKDPEVDERRNLTINCEKAVDLDANAPH
SRAMVYPPNVESSPELPKHITIKLDKGOIIVWIVSPNNDKOKYTLKINHCVPE
QVIAEIRKTRKSMILSSEQLKCYLEYGKQYILKVCDCDEFLKYPLOQYKIRSC
IMLGMPNIMLMAKESLTOLPLDFTMPSRSRISTATPYNGCATKSLMTINSAL
RRIICATVYVNNIRIDIKIYRGTIYHGGEPICDNVTOBPCSNPRNMELSYDM
IDPLPRAARLCLSTICSVGRKGAKEHCPLMNGNINLFDYDTLVSGMALNLAVPH
GLEDLNPIGVYGSNPNKTEPCLELEFDFWFSNPVPEPMTVIEEHANWTISREGFN
SHAGLSNRILARDNELRENDKQRLAICTDPLESEITOEKDFLMSHRHCVTIPETLP
KLLSVKNSRDEVAQMYCLVADMPPIKEQAMELDCNYPDPVRAFAVRLCEKLT
DKLSQYLIQIOLVQVLYKEQYLDNLVFLRLKALTNORIGHEFFHMLKSEHNKTVSO
RGLLESYCRACGMYLKLHNOVEMEKILNTDILKOEKDETOKYOUMEVQOMR
RDEMDALQGLSPINPAHQGNLRLKROVAMEKILNTDILKOEKDETOKYOUMEVQOMR
IIFKNGDDLRODMTLQIIRIMENIMWQGNLDRLMPLGCLSGDCVGLIYVRNSHT
IMQIOCKGGLKGLQFNSHTLHOMLKNKGEITDAILFTRSCAGCVATFIIIG
DHRNSINWKGDLQFHFIDFGLDHRKKKFGYKRRVPEFLVLODFLVIISGAOECT
KTRERFERQEMCYKAYLAIROHANLEINLFSGMLSGMPELOSFDIAYIRKTLALDK
TEOEALFEYMKONDAHHGWTTKMDWIFHTIKOHALN"
BASE COUNT      1028 a      581 c      680 g      918 t
ORIGIN

```

Query Match

Best Local Similarity 100.0%; Score 19; DB 3; Length 3207;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGTTCTTACCATCA 19
|||||
Db 878 AGCGTTCTTACCATCA 860

```

RESULT 2
LOCUS AF001076 3452 bp mRNA VRT 08-JUL-1997
DEFINITION Gallus gallus phosphoinositide 3-kinase catalytic subunit mRNA,
complete cds.
ACCESSION AF001076
NID 92245505
VERSION AF001076.1 GI:2245505
KEYWORDS
SOURCE
ORGANISM chicken.
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 3452)
AUTHORS Chang,H.W., Aoki,M., Fruman,D., Auger,K.R., Bellacosa,A.,
Tischlis,P.N., Cantley,L.C., Roberts,T.M. and Vogt,P.K.
TITLE Transformation of chicken cells by the gene encoding the catalytic
subunit of PI 3-kinase
JOURNAL Science 276 (5320), 1848-1850 (1997)
MEDLINE 97334438
REFERENCE 2 (bases 1 to 3452)
AUTHORS Chang,H.W., Aoki,M. and Vogt,P.K.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
Scrrips Research Institute, 10550 N. Torrey Pines Road, Mail-drop
BCC-239, La Jolla, CA 92037, USA
FEATURES
source
1..3452
/organism="Gallus gallus"
/db_xref="taxon:9031"
/tissue_type="Brain"

```

5'UTR
gene

CDS

```

1..150
151..3354
/feature="c-p3k"
/feature="p3k proto-oncogene"
151..3357
/feature="c-p3k"
/codon_start=1
/product="phosphoinositide 3-kinase catalytic subunit"
/protein_id="AAB62534.1"
/db_xref="PID:92245506"
/db_xref="GI:2245506"
/translation="MPPRSSGELMGILHMPRIIVCELLPNGMIVTECLREATLT
IKHEFKARKPYLIQIODESSYTFVYTOGEAREEFDETERRCDLRLOPFLKVI
EVRGNERELILREIGFAIGMTCFEDVAKDPEVDERRNLTINCEKAVDLDANAPH
SRAMVYPPNVESSPELPKHITIKLDKGOIIVWIVSPNNDKOKYTLKINHCVPE
QVIAEIRKTRKSMILSSEQLKCYLEYGKQYILKVCDCDEFLKYPLOQYKIRSC
IMLGMPNIMLMAKESLTOLPLDFTMPSRSRISTATPYNGCATKSLMTINSAL
RRIICATVYVNNIRIDIKIYRGTIYHGGEPICDNVTOBPCSNPRNMELSYDM
IDPLPRAARLCLSTICSVGRKGAKEHCPLMNGNINLFDYDTLVSGMALNLAVPH
GLEDLNPIGVYGSNPNKTEPCLELEFDFWFSNPVPEPMTVIEEHANWTISREGFN
SHAGLSNRILARDNELRENDKQRLAICTDPLESEITOEKDFLMSHRHCVTIPETLP
KLLSVKNSRDEVAQMYCLVADMPPIKEQAMELDCNYPDPVRAFAVRLCEKLT
DKLSQYLIQIOLVQVLYKEQYLDNLVFLRLKALTNORIGHEFFHMLKSEHNKTVSO
RGLLESYCRACGMYLKLHNOVEMEKILNTDILKOEKDETOKYOUMEVQOMR
RDEMDALQGLSPINPAHQGNLRLKROVAMEKILNTDILKOEKDETOKYOUMEVQOMR
IIFKNGDDLRODMTLQIIRIMENIMWQGNLDRLMPLGCLSGDCVGLIYVRNSHT
IMQIOCKGGLKGLQFNSHTLHOMLKNKGEITDAILFTRSCAGCVATFIIIG
DHRNSINWKGDLQFHFIDFGLDHRKKKFGYKRRVPEFLVLODFLVIISGAOECT
KTRERFERQEMCYKAYLAIROHANLEINLFSGMLSGMPELOSFDIAYIRKTLALDK
TEOEALFEYMKONDAHHGWTTKMDWIFHTIKOHALN"
BASE COUNT      1073 a      672 c      783 g      924 t
ORIGIN

```

Query Match

Best Local Similarity 100.0%; Score 19; DB 4; Length 3452;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCGTTCTTACCATCA 19
|||||
Db 1028 AGCGTTCTTACCATCA 1010

```

RESULT 3
LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
NID 92294345
VERSION A37232.1 GI:2294345
KEYWORDS
SOURCE
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3498)
AUTHORS Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
TITLE EUCHARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 1 17-FEB-1994;
IMP CANCER RES TRCH (GB)
Other publication JP 8503124T 960409.
COMMENT
FEATURES
source
1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/feature="unnamed protein product"
/codon_start=1
/protein_id="CA02284.1"
/db_xref="PID:e304808"
/db_xref="PID:g2294346"
/db_xref="GI:2294346"

```


/translation="MPRPSSGELMGTHIMPRLIVECLPNGMIVTLECREATLIT
IKHELFKARKYPLHOLLODESSYIFVSYQEARREPEFDETRLCRLRLEPPLKAT
EPVGNREKILNREIGFALGMPCEFPVNDPEVDFRRNLINVCRAVDLRDINSPE
SRMAYVPPNVESSPELPKHTYKLDGQIIVYIVVSPNNDKOKYTLKINHCVSE
OVIAEARKTRSMILSSEDLKLCVLEIYQKYLKVCGEDEFLKYSQYKIRSC
IMGRMNLMMAKESYLSOLPMDCFMPSYSRISTATPYMNGESTKSLMWINSAL
RIKLICATYVNNIRDIDKIYVRGTHGGEPLODNNYTORPCSNRNMMNLADYI
IPDLPRARICLSICSVKGRKGAKEEHPPLAMGNINLPDITDILVSGMAMNLMPVPH
GLEDLNPICVGTGSPNPKETPCLELEFDEWSSVVKFPMVSIIEHNAWSRREGFESY
SHAQLSRLARNDELRENDEKOLRAICTRDLSEITDEKDFLMSHRAYCVTIPETILP
KLLTSVWNSRDEVAQMYCLVQWMPPIKPEQAMLLDCNPDPMVRFGEAVCLERKYL
RFGLLSYSCBACGMYLKHNROYEAKELINLIDILKORKEDETQYOKMFLVEQOR
RPFMDLQGLFSLNPAHOLGNLRECKRIMSARKRPLMLNMPNDIMSELHFNONE
IIFRNGDLDQDMITLTIQIRIMENIQONQGLDMLRPLGCLSGDCVGLLEVYNSHT
IMOJOCKGKLGALQFNSHTLHOMLKDNKEEIDDAIDLEFRSCACVATFLIGIG
DRHNSNIMVDDQGLFHDHGFHLDHKKRKGKREVRPEVLTDQFLIVISKAOECT
KTRFERFOEMCYKAYLAIRQHANLFINLSMMLGSGMPELQSPDDIAYIRKTLALDK
TEQALFYPKQMDAHGGWTTKMDWIFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t
ORIGIN

Query Match 100.0%; Score 19; DB 5; Length 3498;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCTTCTTAGCCATCA 19
|||||
Uu 878 AGGCTTCTTAGCCATCA 860

RESULT 4
HSPH13K 3424 bp mRNA PRI 24-AUG-1995
LOCUS H.sapiens mRNA for phosphatidylinositol 3-kinase.
DEFINITION 229090
ACCESSION 9472990
VERSION 229090.1 GI:472990
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 1068)
AUTHORS Volinia,S., Hiles,I., Ormondroyd,E., Nizetic,D., Antonacci,R.,
Rocchi,M. and Waterfield,M.D.
TITLE Molecular cloning, cDNA sequence, and chromosomal localization of
the human phosphatidylinositol 3-kinase p110 alpha (PIK3CA) gene
JOURNAL Genomics 24 (3), 472-477 (1994)
MEDLINE 95229146
REFERENCE 2 (bases 1 to 3424)
AUTHORS Volinia,S.
TITLE Direct Submission
JOURNAL Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK

FEATURES
source
1..3424
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CA82333.1"
/db_xref="PID:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPRPSSGELMGTHIMPRLIVECLPNGMIVTLECREATLIT
IKHELFKARKYPLHOLLODESSYIFVSYQEARREPEFDETRLCRLRLEPPLKAT
EPVGNREKILNREIGFALGMPCEFPVNDPEVDFRRNLINVCRAVDLRDINSPE
SRMAYVPPNVESSPELPKHTYKLDGQIIVYIVVSPNNDKOKYTLKINHCVSE
OVIAEARKTRSMILSSEDLKLCVLEIYQKYLKVCGEDEFLKYSQYKIRSC
IMGRMNLMMAKESYLSOLPMDCFMPSYSRISTATPYMNGESTKSLMWINSAL
RIKLICATYVNNIRDIDKIYVRGTHGGEPLODNNYTORPCSNRNMMNLADYI
IPDLPRARICLSICSVKGRKGAKEEHPPLAMGNINLPDITDILVSGMAMNLMPVPH
GLEDLNPICVGTGSPNPKETPCLELEFDEWSSVVKFPMVSIIEHNAWSRREGFESY
SHAQLSRLARNDELRENDEKOLRAICTRDLSEITDEKDFLMSHRAYCVTIPETILP
KLLTSVWNSRDEVAQMYCLVQWMPPIKPEQAMLLDCNPDPMVRFGEAVCLERKYL
RFGLLSYSCBACGMYLKHNROYEAKELINLIDILKORKEDETQYOKMFLVEQOR
RPFMDLQGLFSLNPAHOLGNLRECKRIMSARKRPLMLNMPNDIMSELHFNONE
IIFRNGDLDQDMITLTIQIRIMENIQONQGLDMLRPLGCLSGDCVGLLEVYNSHT
IMOJOCKGKLGALQFNSHTLHOMLKDNKEEIDDAIDLEFRSCACVATFLIGIG
DRHNSNIMVDDQGLFHDHGFHLDHKKRKGKREVRPEVLTDQFLIVISKAOECT
KTRFERFOEMCYKAYLAIRQHANLFINLSMMLGSGMPELQSPDDIAYIRKTLALDK
TEQALFYPKQMDAHGGWTTKMDWIFHTIKOHALN"

SRMAYVPPNVESSPELPKHTYKLDGQIIVYIVVSPNNDKOKYTLKINHCVSE
OVIAEARKTRSMILSSEDLKLCVLEIYQKYLKVCGEDEFLKYSQYKIRSC
IMGRMNLMMAKESYLSOLPMDCFMPSYSRISTATPYMNGESTKSLMWINSAL
RIKLICATYVNNIRDIDKIYVRGTHGGEPLODNNYTORPCSNRNMMNLADYI
IPDLPRARICLSICSVKGRKGAKEEHPPLAMGNINLPDITDILVSGMAMNLMPVPH
GLEDLNPICVGTGSPNPKETPCLELEFDEWSSVVKFPMVSIIEHNAWSRREGFESY
SHAQLSRLARNDELRENDEKOLRAICTRDLSEITDEKDFLMSHRAYCVTIPETILP
KLLTSVWNSRDEVAQMYCLVQWMPPIKPEQAMLLDCNPDPMVRFGEAVCLERKYL
RFGLLSYSCBACGMYLKHNROYEAKELINLIDILKORKEDETQYOKMFLVEQOR
RPFMDLQGLFSLNPAHOLGNLRECKRIMSARKRPLMLNMPNDIMSELHFNONE
IIFRNGDLDQDMITLTIQIRIMENIQONQGLDMLRPLGCLSGDCVGLLEVYNSHT
IMOJOCKGKLGALQFNSHTLHOMLKDNKEEIDDAIDLEFRSCACVATFLIGIG
DRHNSNIMVDDQGLFHDHGFHLDHKKRKGKREVRPEVLTDQFLIVISKAOECT
KTRFERFOEMCYKAYLAIRQHANLFINLSMMLGSGMPELQSPDDIAYIRKTLALDK
TEQALFYPKQMDAHGGWTTKMDWIFHTIKOHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 100.0%; Score 19; DB 10; Length 3424;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGCTTCTTAGCCATCA 19
|||||
Db 890 AGGCTTCTTAGCCATCA 872

RESULT 5
HSU79143/c 3207 bp mRNA PRI 02-JAN-1997
LOCUS Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA,
DEFINITION complete cds.
ACCESSION U79143
VERSION 94763625
KEYWORDS U79143.1 GI:1763625
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3207)
AUTHORS Sturdivant,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
Olliff,A. and Helmbrook,D.C.
TITLE Catalytic Activity of the p110-alpha Subunit of Human
Phosphoinositide 3'-Hydroxykinase is Required for Signal
Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Sturdivant,S.M., Ahern,J., Conroy,R.R., Barnett,S.F., Ledder,L.M.,
Olliff,A. and Helmbrook,D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
Sumneytown Pike, West Point, PA 19486, USA

FEATURES
source
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..3207
/function="phospholipid kinase"
/codon_start=1
/product="phosphoinositide 3'-hydroxykinase p110-alpha
subunit"
/protein_id="AA839753.1"
/db_xref="PID:91763626"
/db_xref="GI:1763626"
/translation="MPRPSSGELMGTHIMPRLIVECLPNGMIVTLECREATLIT
IKHELFKARKYPLHOLLODESSYIFVSYQEARREPEFDETRLCRLRLEPPLKAT
EPVGNREKILNREIGFALGMPCEFPVNDPEVDFRRNLINVCRAVDLRDINSPE
SRMAYVPPNVESSPELPKHTYKLDGQIIVYIVVSPNNDKOKYTLKINHCVSE
OVIAEARKTRSMILSSEDLKLCVLEIYQKYLKVCGEDEFLKYSQYKIRSC
IMGRMNLMMAKESYLSOLPMDCFMPSYSRISTATPYMNGESTKSLMWINSAL
RIKLICATYVNNIRDIDKIYVRGTHGGEPLODNNYTORPCSNRNMMNLADYI
IPDLPRARICLSICSVKGRKGAKEEHPPLAMGNINLPDITDILVSGMAMNLMPVPH
GLEDLNPICVGTGSPNPKETPCLELEFDEWSSVVKFPMVSIIEHNAWSRREGFESY
SHAQLSRLARNDELRENDEKOLRAICTRDLSEITDEKDFLMSHRAYCVTIPETILP
KLLTSVWNSRDEVAQMYCLVQWMPPIKPEQAMLLDCNPDPMVRFGEAVCLERKYL
RFGLLSYSCBACGMYLKHNROYEAKELINLIDILKORKEDETQYOKMFLVEQOR
RPFMDLQGLFSLNPAHOLGNLRECKRIMSARKRPLMLNMPNDIMSELHFNONE
IIFRNGDLDQDMITLTIQIRIMENIQONQGLDMLRPLGCLSGDCVGLLEVYNSHT
IMOJOCKGKLGALQFNSHTLHOMLKDNKEEIDDAIDLEFRSCACVATFLIGIG
DRHNSNIMVDDQGLFHDHGFHLDHKKRKGKREVRPEVLTDQFLIVISKAOECT
KTRFERFOEMCYKAYLAIRQHANLFINLSMMLGSGMPELQSPDDIAYIRKTLALDK
TEQALFYPKQMDAHGGWTTKMDWIFHTIKOHALN"

```

RILKATVYVNRIDIDKIYVRCIYHGEPLCDNVNTQVPCSNPRNWNLTNDI
IPDPRAKRLCLSTSVKRGKAGKEBEPCLMGNINLDTIDTLVSGMALNLPVPH
GLEDLNLIVGTGSPNKEETPCLELEFDMFSSVFKPDMASVIEBHANNVSREGFSY
SHAGLSNRLADNELRENDKEQLAISTRDLSLETBEKEDLMSHRAYCVTIEILP
KLLSVKNSRDEVAOMYCLVMDPPIKPEOAMELDCNPPDMVGRGAVACLEIYLT
DKLSOYLIOLOVYAKYEOYLDNLVRLKRLKALNORIGHEFFMHLKSEMHNTVSQ
RFGLLSEYCRACGMYLKLHNOYVAMEKILNLDILKOEKDETOXYOMFVYQMR
RPPMDALOGFLSPNPAHQGNLRECRIMSSAKRPLMWNENPDIMSELIRONNE
IIFKNGDDLRODMTLQIIRIMENIMWONGDLRLPYGLSTIGCVGLIEVNRSHI
IMOIOCKGGLGALQFNSHTLHOMLKNKEIYDAIDLETRSCAGYCAVFTILIGIS
DRHNSNIMVKDQGLFHDGFHLDHKKKKEKRYEPVLTODFLIVISKAQECT
KTREPEROECMAYALAIROHANFLNLFMSMLGSGMPELQSFODIAYIKRTALDK
TEGALEEFMKOMDANHGGWTKMDWIFHTIKOHALN"

BASE COUNT      1043 a      586 c      670 g      908 t
ORIGIN

Query Match      100.0%; Score 19; DB 10; Length 3207;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGGCTTCTTAGCCATCA 19
      ||||||||||||||||
Db      878 AGGCTTCTTAGCCATCA 860

RESULT 6
AF001075/c      3389 bp      mRNA      VRL      08-JUL-1997
LOCUS      Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
DEFINITION      subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
partial cds.
ACCESSION      AF001075
NID      92245502
VERSION      AF001075.1 GI:2245502
KEYWORDS      Avian sarcoma virus 16.
SOURCE      Avian sarcoma virus 16.
ORGANISM      Avian sarcoma virus 16.
REFERENCE      1 (bases 1 to 3389)
AUTHORS      Chang, H.W., Aoki, M., Fruman, D., Auger, K.R., Bellacosa, A.,
Tsichlis, P.N., Cantley, L.C., Roberts, T.M., and Vogt, P.K.
TITLE      Transformation of chicken cells by the gene encoding the catalytic
subunit of pi 3-kinase
JOURNAL      Science 276 (5320), 1848-1850 (1997)
MEDLINE      97334438
REFERENCE      2 (bases 1 to 3389)
AUTHORS      Chang, H.W., Aoki, M. and Vogt, P.K.
TITLE      Direct Submission
JOURNAL      Submitted (23-APR-1997) Molecular & Experimental Medicine, The
Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
BCC-239, La Jolla, CA 92037, USA
FEATURES
source      1. 3389
/organism="Avian sarcoma virus 16"
/db_xref="taxon:60629"
<1..33252
/gene="gag-v-p3k"
/note="fusion gene of gag and v-p3k"
1..78
/gene="gag-v-p3k"
/note="derived from gag gene"
<1..33252
/gene="gag-v-p3k"
/codon_start=1
/product="gag-v-phosphoinositide 3-kinase catalytic
subunit fusion protein"
/protein_id="AAB6533.1"
/db_xref="PID:g2245503"
/db_xref="GI:2245503"
/translat="QOLIRFAPSTLTPGELIYKVLDRKASHMLPRILVECLPVG
MITVECLREATLTIKHELFKARKAYPLVLDDESYIFSVTQAEAREFEDETR
RLCDRLFQPLKVIKIEPVGNEKILNKIGFALIGMPCIEDMVKDPVDFRNIIIN

RILKATVYVNRIDIDKIYVRCIYHGEPLCDNVNTQVPCSNPRNWNLTNDI
KQRTTLKINHOCVEQVYIARIRKTKSMILSSQLKLVLEVOGKTLVVCGDETL
LEKPLSQKQYIRSCINMIGRPNIMLMAKSLVYQLDPLDVELEYSRSTIARVYN
GEARLKLWTLNLSLRIRILCAIVVNIIDIKIYRTGIYHGEPLCDNVNTQVYP
CSNRMENLMSYDIEDLIDLPRAALSTSVKRGKAGKEBEPCLMGNINLDPVPH
LVSCKMNLMAVPHGLEDLNLIVGTGSPNKEETPCLELEFDMFSSVFKPDMASV
EHAWTISREIGFVSTAGLSNRLADNELRENDKEQLAISTRDLSLETBEKEDL
WSHRHVCVKTPELIPKLLSVKNSRDEVAOMYCLVMDPPIKPEOAMELDCNPPD
MYRFAVAVCRLEKYLTDKLSOYLIOLOVYAKYEOYLDNLVRLKRLKALNORIGHE
FVHLKSKMHNTVSQORGLILEYCRACGMYLKLHNOYVAMEKILNLDILKOEKED
ETGVOKMFLVEOVRPDMFALOGFLSPNPAHQGNLRECRIMSSAKRPLMWNEN
ENPMIEMELFONNELIIFKNGDDLRODMTLQIIRIMENIMWONGDLRLPYGLST
GDCVGLIEVNRSHITIMOIOCKGGLGALQFNSHTLHOMLKNKEIYDAIDLETR
SCAGYCAVFTILIGDRHNSNIMVKDQGLFHDGFHLDHKKKKEKRYEPVLTODFL
ODLIVISKAQECTKTREPEROECMAYALAIROHANFLNLFMSMLGSGMPELQSF
FDDIAYIRKTLALDKTEGALEEFMKOMDANHGGWTKMDWIFHTIKOHALN"

BASE COUNT      1071 a      655 c      751 g      912 t
ORIGIN

Query Match      100.0%; Score 19; DB 17; Length 3389;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGGCTTCTTAGCCATCA 19
      ||||||||||||||||
Db      923 AGGCTTCTTAGCCATCA 905

RESULT 7
MM003279/c      3207 bp      mRNA      ROD      25-MAY-1994
LOCUS      Mus musculus Balb/c phosphatidylinositol 3-kinase 110 kDa subunit
DEFINITION      mRNA, complete cds.
ACCESSION      9414994
NID      003279
VERSION      003279.1 GI:414994
KEYWORDS      house mouse.
SOURCE      Mus musculus.
ORGANISM      Mus musculus.
REFERENCE      1 (bases 1 to 3207)
AUTHORS      Klippel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
TITLE      The interaction of small domains between the subunits of
phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL      Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE      94187738
REFERENCE      2 (bases 1 to 3207)
AUTHORS      Klippel-Giese, A.
TITLE      Direct Submission
JOURNAL      Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Daichi
Research Center, University of California San Francisco, 505
Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
FEATURES
source      1. 3207
/organism="Mus musculus"
/db_xref="taxon:10090"

```

CDS

1. 3207
/standard_name="p110 protein"
/codon_start=1
/product="phosphatidylinositol 3-kinase 110 kDa subunit"
/protein_id="AA18334.1"
/db_xref="pid:9414995"
/db_xref="GI:414995"
/translation="MPPRSSGEIMGHMPPRLVBCLEPNGMIVTECREATLVY
IKHLEFRAKRYPLHQLQDETSTYFVSVQEAREREFEEDTRRLCDRLRQPLKVI
EPVGNREKILNREIGVIGMPVCEFDMDVNDPEFRNINVCKEAVDRLDPSH
SRMYVVPNVESSPELPKHIYNKLDKQIIVLVIVIPNNDKQKYLKINHCVPE
QVIAERKTRKSMLESEOLKCVLEQGYIIVKVCDCDEPFEKYSQYKIRSC
IMIGRPNIMAKRESLYSOLPDSFTMPSRSTSTATPYPMNETSKRSIMVNSAL
RIKICATYVAVNRDIDKITYRGITGHGGPCLDNNVTQVPSNPNMMLVTDIT
IPDLPLRLRLCLTSCYKGRGAEHCPLMNGINLEFYDTLVSGKMLNLMPVPR
GLDLNLPLVGTGSPNKEPCLLEFDMFSSVVFPEFVSNVYEEHANSVREAGFSY
SHTGSLNLRDNLRENDKQLRALCRDPLETTEQKDFLMSHRYCYTIEILP
KLTSVKNWSRDEVAOMCYKMDPPIKPEQAMELDCNPDMPVRSFAVRCLEKYL
DDKLSOYLIOVYLYKBOYDNLVPRLLKALTNORIGHFPMHLKSEHNTVSO
RFGILLSEYCPACGMYLKHLNROYBAEKILNLDILQEKDETOKQKFLVEOKR
QPDMDALQGLSPNPAHQNLRLSECRIMSSAKRPLMNMENPDIMSELQNMNE
IIFNNGDRLQMDLTLQIIRIMENIMONOGDLMLPYGCLISGCVGLIEVNSHT
IMJOICKGGLKGALEFNSHTLHOMLKDNKEIGDAIDLEFSCAGCAVATFLLIG
DRHNSNIMVNDGOLFHDPEGHFLDHKKKKRGRYREYFVLTDDELIVISKGOEY
KTRFEFERQECYKAYLAIRHNLPLFLESMIGSSMPELQSFDDIAYIKITLALDK
TEQALEFTKQMDAHHGWTIMDMVLEHTIKOHALN"

BASE COUNT 958 a 693 c 734 g 822 t

ORIGIN

Query Match 91.6% Score 17.4: DB 12: Length 3207;
Best Local Similarity 94.7%: Fred. No. 26;
Matches 18: Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 AGGCTTCTTACCATCA 19
|||||||
Db 878 AGGCTTCTTGGCATCA 860

RESULT 8
HS1068E13/c
LOCUS HS1068E13 161222 bp DNA HTG 11-JUN-1999
DEFINITION Homo sapiens chromosome 20 clone 1068E13, WORKING DRAFT SEQUENCE,
in unordered pieces.
ACCESSION AL035563
NID 95051828
VERSION AL035563.15 GI:5051828
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
Bates, K.
1 (bases 1 to 161222)
Direct Submission
Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquey@anger.ac.uk Clone requests: clonerequest@anger.ac.uk
On Jun 12, 1999 this sequence version replaced gi:4902588.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj1068E13 Contig_ID: 00053 acc-AL035563
Length: 1949 bp Unfinished: dj1068E13 Contig_ID: 00889
acc-AL035563 Length: 1412 bp Unfinished: dj1068E13 Contig_ID:
02208 acc-AL035563 Length: 1145 bp Unfinished: dj1068E13
Contig_ID: 01831 acc-AL035563 Length: 1105 bp Unfinished:
dj1068E13 Contig_ID: 00957 acc-AL035563 Length: 148176 bp
Unfinished: dj1068E13 Contig_ID: 02504 acc-AL035563 Length: 1333
bp Unfinished: dj1068E13 Contig_ID: 02342 acc-AL035563 Length: 1333

COMMENT

1302 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source
Location/Qualifiers
1. 161222
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="1068E13"

BASE COUNT 48587 a 33616 c 32167 g 42049 t 4803 others

ORIGIN

Query Match 84.2% Score 16: DB 34: Length 161222;
Best Local Similarity 100.0%: Pred. No. 1.2e+02;
Matches 16: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 GCTTCTTACCATC 18
|||||||
Db 44809 GCTTCTTACCATC 44794

RESULT 9
AE001605
LOCUS AE001605 10553 bp DNA BCT 22-FEB-1999
DEFINITION Chlamydia pneumoniae section 21 of 103 of the complete genome.
ACCESSION AE001605 AE001363
NID 94376455
VERSION AE001605.1 GI:4376455
KEYWORDS
SOURCE Chlamydia pneumoniae.
ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 10553)
AUTHORS Kalman, S., Mitchell, R.W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 10553)
AUTHORS Kalman, S., Mitchell, R., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
Source
Location/Qualifiers
1. 10553
/organism="Chlamydia pneumoniae"
/strain="CWL029"
/db_xref="taxon:83558"
255. 1427
/gene="Cpn0186"
255. 1427
/gene="Cpn0186"
/transl_table=1
/codon_start=1
/product="similarity to Ctl19 Inca"
/protein_id="AA018339.1"
/db_xref="PID:94376456"
/db_xref="GI:4376456"
/translation="MSPPVNTPSAPNIPIPAPPTTGPITPRSSFEIKYIYAKYI
LFAIATASGALITIGLIGALTPGIALIYIFVSWLGLILKDISGGEERLRE
EVSRTENORLVTITLLEVEVDLKAADOLLEAFENNGINTKTAEDLEEDV
SKLSEQLAERINQLOIQAANGDOETSEKTLISGMDSKVVOINTSIOALKVLG
QEWQDQATYKNAQEOIQALQAIILGNHNSYALQKSYENLVLQDALTFRVVELE
SENLSQACSAKREIEKLAQHEYSIDQIDAMLAQONLAEVYALEKMQDAQKAE
SEFIACVADRTRGFRRETPPTTPEVGEDESEDEGGTPPVSPSSPVDRATGGO"

gene
1604. 2404
/gene="Cpn0187"
/gene="Cpn0187"
/codon_start=1

/transl_table-11
/product- "predicted methylase"
/protein_id- "AAD18340.1"
/db_xref- "PID:94376457"
/db_xref- "GI:4376457"
/translation- "MHSKPLSRKRNKSSHEKEETSPDCLASSTYKRIYODKHYYHRETI
LROLLSLTSGSKSVLDICGGGFLERAPRRCRIIGIDISSRLATRAKMAKSNVSH
QERIVADSKRIEVEPTLFSHAVALSLQWMEPEGHINTATLLEPLGOFYVLMHP
CFRIIPRASSWHYDENKRAISRHIDRYLSPKIPIMAPGQDPSLSEFHPSTYWEK
ELSSHGLVLSGLEWTSKSTGTCKRAKEMICRKEPFLMISCIRIK"
2467.3759
/gene- "Cp00188"
2467.3759
/gene- "Cp00188"
/codon_start-1
/transl_table-11
/product- "Cti32 hypothetical protein"
/protein_id- "AAD18341.1"
/db_xref- "PID:94376458"
/db_xref- "GI:4376458"
/translation- "MFRKLFPSKRRKTGQRKRLNNGLLQAIQISIKVLLHNASKREA
CVLSYGLTVCPIVLEFLRSQHEFTNLNKKEMLIKEPDYKRPVAVAEAYHATE
SNIGLVGSFEFCWAGIIMLSLEDGLMKIRTSWPTSLKRLVSYETLVSAMI
FIIVCGSMIYITQIMPIQYAKFLPSHSMATLFSRFPYLLYLALFCCYAFRLV
AIOKTSALSTLIGSVMIYFQKAFPSLOYSTNSATSGTALVALPSLLIYITMI
YLEGCAITFIQNRCTFIPLGDKIIPSCITQITSTYITLATTROPNEGLSLTQF
IAKQSVPIGEVSCUDVLEKEGFLPPYNGYQVFNFSELTIKDIAADKLHREIFRK
FNPDLGITFENFONIFNOASKNKENTLSEIARRIK"
3756.7184
/gene- "Cp00189"
3756.7184
/gene- "Cp00189"
/codon_start-1
/transl_table-11
/product- "Cti31 homolog (Possible Transmembrane Protein)"
/protein_id- "AAD18342.1"
/db_xref- "PID:94376459"
/db_xref- "GI:4376459"
/translation- "MKRRWKLKLGICLSSYVGLFLLPQLSTESKRYVFSLIH
KESGLSCSAEELKISWFGQTARKIKLGEAKDEVSAEFELDGLRLILYKRRKG
ITLWSGLKINEPASIDHPSVSHLDPSLTYLMDCKIISEHGFITMKTVSGSLSVS
GYFLKESSEKFMTCVSEDOQGNIFIESVLSYSAOFSSVPAFFKPIIAPF
MDHLSEYEDILINSAETHNDGKISMTAGEGNOIOMKLGHIKSTFYVSGSSSF
IELPELASALCNOIIPISPTSKOIHATYSARAKPIDITKRRKHIEITSQALPVA
IHKPDNLAQLADTKLGIKTEKESDIRSSSTVYAGSPSHKGLISIDNKHILIK
FRLOAQLPHYTLRALFPOPEVINVPLDAVYSLNLEGYKNAHLEADAILDPKFKL
SCMSGTMRNLEFGQGYHFENKQOEIILSPHSYAEARPSGAQITDNLFFPKFG
KITARENELLIHAKFGSPNEPIKETTSILIHGQFCSLPLSVSNHAEFHLKTFES
FHTDGGFVTKGMLQALIENTPOYDLNNTIRILIDLLSLDESSTSPSKDLIOGSG
EITSPLDSTITKTKYQVRLSPYSGSGLNFPVNNRPDQNKLTILSNFKSALIGE
LKIYDPFMRKLSGSGTGTQWVEYSPERIASFFKNASCSPCLLHRTANRDLISLSC
PEETKGLSCTLLAAGLEGSLAETAPLFIYDANKETFIINDKGLRANNDLAKLEY
DLKQSLAPRODSKTLAEFLSEGOVDHLESPESREKFKQAMLIHPSSEIAGIIPSP
GLKAOISLAPRINIVSINKNAFERGEPDIDMSQALQIILINSEKSLIRENLT
AHLISINEDVKAIOEPNPLAGGAYSOVPTLEIDKONFYLPRIYSEEPFOAT
LDMGKSIANTGTMVALROFLDITDOKKOSVETFIIFSQVGSITCKRDLADIR
IRIALMGKTIDIANRDLFMTLDIDPEVITKTFHSTSLKTNFLLIKIRGSIISFEVMS
SAVARIALKSTSYLGNPFSSLADKLFSSLDGSTPPPTVHPFMEKSNFDSIENK"
7275.8633
/gene- "Cp00190"
7275.8633
/gene- "Cp00190"
/codon_start-1
/transl_table-11
/product- "hypothetical protein"
/protein_id- "AAD18343.1"
/db_xref- "PID:94376460"
/db_xref- "GI:4376460"

SPPLPGLTSGISGFTWKSASKNSGSPFSAKLRKTESDTCQITSTISGQAG
TTWLSLKVIVPSIFQIEKPEVQSLVSYEDWLPIDINIFMSQPRITPLALQOTM
LAGOKYDILDELAAHQINOTLMTISPCNSRSLQKOTNOFENSPVDYIVHAHSCHWS
GF"
/complement(8667.9347)
/gene- "glnp"
/note- "Cp00191"
/complement(8667.9347)
/gene- "glnp"
/codon_start-1
/transl_table-11
/product- "ABC Amino Acid Transporter ATPase"
/protein_id- "AAD18344.1"
/db_xref- "PID:94376461"
/db_xref- "GI:4376461"
/translation- "MTIRYRNLAYSVNKKIIDGVTFSLERHITLTFVKGSGSGKMTI
LKAALGVOPGTGDIWIEGEAPALVOGELEFSHMTVGNCTHPQIHIGRSTEAR
KAPELLHLDIEVAKNKPDSLGQOKORVATVRSICMDKHTLLFDEPTSAIDPRATA
SRHLELTIRDELIVGLTTHMORVHSCLDRIYIIDOGTVAGYIDKRDGELDSGHL
SKYHSAQ"
/complement(9337.9990)
/gene- "glnp"
/note- "Cp00192"
/complement(9337.9990)
/gene- "glnp"
/codon_start-1
/transl_table-11
/product- "ABC Amino Acid Transporter Permease"
/protein_id- "AAD18345.1"
/db_xref- "PID:94376462"
/db_xref- "GI:4376462"
/translation- "MDHWLAIARLLRRCGYTLVCYSGIGILGSLIGLIGVTSYLF
FSKTLRLANSIVIVIKGTPLFTQILIIYFGLPEVLPIEPTPLVAGIILASVNSAYL
AENIRGINSLSIQGESAMVLYGKRYOIFVYIYPOVEKNLPSLNEPVSLSIES
ILMVGVPELTAVTRDIVERSREINPMENLICAGLFTMTSPSCISRLSEKRSYDN"
/complement(10039.10482)
/gene- "argr"
/note- "Cp00193"
/complement(10039.10482)
/gene- "argr"
/codon_start-1
/transl_table-11
/product- "Arginine Repressor"
/protein_id- "AAD18346.1"
/db_xref- "PID:94376463"
/db_xref- "GI:4376463"
/translation- "MKKRVTTIDEALKEILLREGAATQEBLAKLLAOGFATQSSVS
RKRIOAVVAGEGRGARYSLPSTETKTRHLYVLSIRHNASLIVRTVPGSASMTAAL
LDQGLKDELITAGDDPIFVPRIDEGSLPLMWSIANLVLQVFD"

Query Match 83.2% Score 15.8; DB 2; Length 1053;
Best local similarity 89.5%; Pred. No. 1.7e-02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
DB 3188 AGGCATTCTTTAGCCTTCA 3206
RESULT 10
LNU07640/c 1138 bp DNA BCT 05-JAN-1995
LOCUS LNU07640 Lactococcus lactis triosephosphate isomerase (tpl) gene, complete
DEFINITION cds
ACCESSION 007640
NID 9538127
VERSION 007640.1 GI:538127
KEYWORDS Lactococcus lactis
SOURCE Lactococcus lactis
ORGANISM Lactococcus lactis

Bacteria: Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
Lactococcus.
REFERENCE
1 (bases 1 to 1138)
AUTHORS
Cancilla,M.R., Davidson,B.E., Hallier,A.J., Nguyen,N.Y. and Thompson,J.
TITLE
The Lactococcus lactis triosephosphate isomerase gene, tpi, is monocistronic
JOURNAL
Microbiology 141 (Pt 1), 229-238 (1995)
MEDLINE
95202084
2 (bases 1 to 1138)
AUTHORS
Davidson,B.E.
TITLE
Direct Submission
JOURNAL
Submitted (10-MAR-1994) Barrie E. Davidson, University of Melbourne, Biochemistry, Parkville, Victoria, 3052, Australia
COMMENT
On Sep 15, 1994 this sequence version replaced g1:537285.
FEATURES
source
location/Qualifiers
1..1138
/organism="Lactococcus lactis"
/sub_species="lactis"
/isolate="LM0230"
/db_xref="taxon:1358"
/clone_id="lambda GEM11"
/map="1.818 to 1.978 Mbp"
72..77
-35_signal
-10_signal
102..107
115..988
mrna
/evidence="experimental"
165..170
RBS
178..936
gene
/gene="tpi"
178..936
CDS
/gene="tpi"
/EC_number="5.3.1.1"
/codon_start=1
/transl_table=1
/evidence="experimental"
/product="triosephosphate isomerase"
/protein_id="AAC43268.1"
/db_xref="PID:g1:537286"
/db_xref="GI:537286"
/translation="MSRRPITAGNMKNTLSEAQFVAVKNNLPSSDNVESYIGAP
ALFLAPAYLRGSEIKLAENSYFENAGFTGNSPAIYDIDGIEYIIIGHSRRREY
FHEHDEDINKKAKAIFPAAGATPICCGSETLETFRAGKTAENWSCQIEAGLTADEY
SNLVAIEPNAIGTGTATNEIADETGGVRSYVEKLYGKVESEAVRIQYGVSKPPE
TIEGLMAKENIDGALVGASLEADSFALLEMYR"
958..985
repeat_region
/note="rho-independent terminator"
/function="putative transcription terminator"
/rpt_type="inverted"
/rpt_unit="958..970"
terminator
BASE COUNT 379 a 192 c 235 g 332 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 2; Length 1138;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AGGCTTCTTAGCCATCA 19
1138
Db 863 ATGTTCTTTAGCCATCA 845

RESULT 11
LOCUS D86747 3060 bp mRNA VRT 07-FEB-1999
DEFINITION Chicken nei mRNA for Mr 93 k protein (with EGF-like repeats),
complete cds.
ACCESSION D86747 DA5365
MID 91483183
VERSION D86747.1 GI:1483183
KEYWORDS Mr 93 k protein; nei.

SOURCE
ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Archosauria; Aves;
Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE
1 (bases 1 to 3060)
AUTHORS
Matsunashi,S.
TITLE
Direct Submission
JOURNAL
Submitted (26-JAN-1995) to the DDBJ/EMBL/GenBank databases. Sachiko
Matsunashi, Saga Medical School, Department of Biochemistry, 5-1-1
Nabeshima, Saga, Saga 849, Japan
(E-mail:matsunash@post.saga-med.ac.jp, Tel:0952-31-6511(ex.2265),
Fax:0952-33-2511)
2 (sites)
REFERENCE
Matsunashi,S.
AUTHORS
Unpublished (1996)
JOURNAL
3 (bases 1 to 3030)
REFERENCE
Matsunashi,S., Noji,S., Koyama,E., Myokai,F., Ohuchi,H.,
AUTHORS
Taniguchi,S. and Horii,K.
TITLE
New gene, nei, encoding a M(r) 93 k protein with EGF-like repeats
is strongly expressed in neural tissues of early stage chick
embryos
JOURNAL
Dev. Dyn. 203 (2), 212-222 (1995)
MEDLINE
95383734
COMMENT
On Aug 8, 1996 this sequence version replaced g1:642136.
D45365:Submitted(26-Jan-95) to DDBJ/EMBL/GenBank databases.
LOCATION/Qualifiers
source
1..3060
/organism="Gallus gallus"
/db_xref="taxon:9031"
/dev_stage="9 day old embryo"
118..2568
gene
/gene="nei"
118..2568
CDS
/gene="nei"
/note="with EGF-like repeats"
/codon_start=1
/product="Mr 93 k protein"
/protein_id="BAA13167.1"
/db_xref="PID:g1:642136"
/db_xref="PID:g1:483184"
/db_xref="GI:1483184"
/translation="MESGCGICGTCILICGPGVYRGVDSIQIDVLSGIDPBYAG
VROVPLGNSKAPLFPDTRSVASPELMEIFOKLRNKEFTITVTLKQAHNSGV
IFSLHLDHRYLESSGHRNEIRLHRTGSHRSTVEFVPLADDDWHRLSLAIS
HLIHLVDCNKIYERVENKPEPMDLPGVTFMVGQNNHAGYFKGMDOVLYLMQFGL
SOCEDLNRCPTCDNFGLVOKIMELDIDIAKTSAKSOAEORNRKIKDOICVETCM
KGMVREPESTWDCGKNCCTGNGVOCBALICSIDCPNSALSYVGKKCKEQQSVC
IFESRTYFEGOREYVSSSGCVLFECKDHQMORIPRDSCTILNCPESQOIRLSHSC
KICGHDPCFBGHCMEHSVCRNLDRAVCSCRDGFALREDNAYCEDEVCAGORH
CRENTMGNVNTPGSPMCICIKTGYIRIDYSCHEDECVYTNHCDEMLCFEYVAGHNC
VCKLGYNGVTPVCAKFCGDCRNGAGIASVACAPGFTGSPCEETIDIECSDFVVC
DSRANGLNLPGMVYCECRDGYHDNGFSPSGESCEIDECATGHRSCANDVCPNLNG
GYDCRPHGKNCITGDCIHEDKIRKNGOIVYENRGVSGSGGVCMRRVACDEMP
TYDFCCPECDPRRLSSQCLHDSGLSYNSGWSYQNCQQCCLOGEYDQNPPEPYD
CEESVLPENBCCPRCVTDPCQADTIRNDITKTCIDETNVAVFTSSWIKHGTCTLCO
CKNGHVCSSVDPQLOEL"

Query Match 83.2%; Score 15.8; DB 4; Length 3060;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 AGGCTTCTTAGCCATCA 19
3060
Db 571 AGGCTTCTTAGCCATCA 589

RESULT 12
LOCUS GGFMYLCK/C 7631 bp mRNA VRT 17-FEB-1997
DEFINITION G.gallus mRNA for fibroblast myosin light chain kinase.


```

/citation-[5]
/replace="c"
5037
/citation-[5]
/replace="c"
5248..5721
/codon_start=1
/product="KRP/Relokin"
/protein_id="CAA37058.1"
/db_xref="PDB:9992995"
/db_xref="SMISS-PROT:P11799"
/translation="MAMISGSGRAGSSPTSPINDKVENEDAFLEVAERPHV
PYFKTLIDMEVVGSSARFPDCKTEGYDPPVMMKDDQPKESRKHIDYDEBNC
LTISEVCGDDAKKATCTCKAYVNSLGEATCTAELLVETMKGEGSGEGDEDEEBEE"
BASE COUNT      2435 a      1569 c      1793 g      1834 t
ORIGIN
Query Match      83.2%; Score 15.8; DB 4; Length 7631;
Best Local Similarity 89.5%; Pred. No. 1.8e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0
QY      1 AGGCTTCTCTTAGCCATCA 19
      ||| ||| ||| ||| |||
Db      1474 AGGCTTCTCTTACCATA 1456

RESULT 13
GMLCKR08      253 bp      DNA      VRT      08-AUG-1998
LOCUS      Gallus gallus smooth muscle/non-muscle myosin light chain kinase
DEFINITION      gene, exon 8.
ACCESSION      AF045262
MID      93403177
VERSION      AF045262.1 GI:3403177
KEYWORDS
SEGMENT
SOURCE
ORGANISM
      * 8 of 31
      chicken.
      Gallus gallus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
      Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
      1 (bases 1 to 253)
      Watterson,D.M., Collinge,M., Lukas,T.J., Van Eldik,L.J.,
      Bitukov,K.G., Stepanova,O.V. and Shirinsky,V.P.
      Multiple gene products are produced from a novel protein kinase
      transcription region
      FEBS Lett. 373 (3), 217-220 (1995)
      96033976
      2 (bases 1 to 253)
      Bitukov,K.G., Schavocky,J.P., Shirinsky,V.P., Chibalina,M.V., Van
      Eldik,L.J. and Watterson,D.M.
      Organization of the genetic locus for chicken myosin light chain
      kinase is complex: multiple proteins are encoded and exhibit
      differential expression and localization
      J. Cell. Biochem. 70 (3), 402-413 (1998)
      98370374
      3 (bases 1 to 253)
      Bitukov,K.G., Schavocky,J.P., Shirinsky,V.P., Van Eldik,L.J. and
      Watterson,D.M.
      Direct Submission
      Submitted (30-JAN-1998) Molecular Pharmacology and Biological
      Chemistry, Northwestern University Medical School, 303 E. Chicago
      Ave. Ward-8-196, Chicago, IL 60611, USA
      Location/Qualifiers
      1..253
      /organism="Gallus gallus"
      /db_xref="taxon:9031"
      /tisue_type="liver"
      /dev_stage="adult"
      27..230
      /number=8
FEATURES
source
exon
BASE COUNT      80 a      46 c      50 g      67 t

```

Query Match	83.28;	Score 15.8;	DB 4;	Length 253;
Best Local Similarity	89.5%;	Pred. No. 2.1e+02;		
Matches 17/	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	AGGCTTCTTACGCATCA	19	
Db	146	AGGCTTCTTACCATTA	128	
RESULT	14			
LOCUS	AC007258/c			
DEFINITION	Arabidopsis thaliana chromosome I BAC F23H11 genomic sequence, complete sequence.			
ACCESSION	AC007258			
NID	94733952			
VERSION	AC007258.3			
KEYWORDS	GI:4733952			
ORGANISM	Arabidopsis thaliana			
SOURCE	Arabidopsis thaliana			
REFERENCE	1 (bases 1 to 144422)			
AUTHORS	Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.			
JOURNAL	Submitted (08-APR-1999)			
REFERENCE	2 (bases 1 to 144422)			
AUTHORS	Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.			
JOURNAL	Submitted (04-MAY-1999)			
REFERENCE	3 (bases 1 to 144422)			
AUTHORS	Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.			
JOURNAL	Submitted (16-JUN-1999)			
REFERENCE	4 (bases 1 to 144422)			
AUTHORS	Federici, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altati, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremenetskaia, I., Kim, C., Lenz, C., Li, J., Liu, S., Luros, S., Schwartz, J., Shinn, P., Tortum, M., Vyotskaia, V., Walker, M., Yu, G., Ecker, J., Theologis, A. and Davis, R.W.			
JOURNAL	Submitted (16-JUN-1999)			
COMMENT	<p>On May 4, 1999 this sequence version replaced gi:4678191.</p> <p>Bases 138041-144421 'IGF' Clone F23H11 overlap with bases 1-6382 of 'TAMU' Clone T2K10, Accession: AC005966.</p> <p>e-mail for correspondence: arabseqsequence.stanford.edu</p> <p>Genes with similarity to proteins in the databases are described as 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on genome predict ion software are described as 'hypothetical proteins'.</p>			

The software programs used to predict genes include: Graft (Informatics Group, Oak Ridge National Laboratory, <http://compbio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genome.stanford.edu/~chris/GENSCANW.html>), Fexa (V. Solovjev & A. Salamov, Sang er Centre, <http://genome.sanger.ac.uk/>), NeplantGene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/NeplantGene.html>) and eMocif (Newell Manning, C.G. Wu, T.D. & Brutlag, D.L., <http://mocif.stanford.edu/projects.html>).

FEATURES

```

source
1..144422
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/clone="F23H11"
/chromosome="1"
2303..3607
/gene="F23H11.1"
join(2303..2629,3230..3607)
/gene="F23H11.1"
/note="Similar to glutathione transferase"
/codon_start=1
/protein_id="AAD39312.1"
/db_xref="PID:95080803"
/db_xref="GI:5080803"
/translation="MGKEKVKLGWVSPYAIRPKIALRLKSDYDYVEERLQESKS
ELLKSNFVHKVPYLHNKRPVLSLIVETIDTNNSSAPSLIPSPIDALRWM
SDFVDMKFPALRMAIITKSDAKAKAEVEEGLQLEDAFVLSISKRPFGGAIG
FMIDCFGSFVLLKAREKFKAKLIDESTKPSLCKWADRFLSDETKVAVEIEKVAE
FLQLEVRAQSNASRS"
4467..5517
/gene="F23H11.2"
join(4467..5051,5122..5416,5495..5517)
/gene="F23H11.2"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39313.1"
/db_xref="PID:95080804"
/db_xref="GI:5080804"
/translation="WEIFOKAKAVALLSHHDXYLVADDEESVTQERNSGAAMKWTY
EIRGSTNLIRLKSAGKYVLTASNKPIFGATGKVLQTKPSRLDSSILAMPISDSAL
VKLTRYNGFLRNGSLPMPRNSVTHDIPHSATQEVYLMHIDVVEILPANSOSHH
OHLQLOPSPSLHSDSLDFTPGSPSRSDREFRQSTDSLAVGSPKSGEVITYHVA
DDDDVEDDSVEVSFTKNGVEELTRKLEBSNVEVYICTRSLNGKLPFLRLQL
PPNADMTVVLVPPSSKIAEKFTK"
complement(5712..8207)
/gene="F23H11.3"
complement(join(5712..7462,7970..8039,8181..8207))
/gene="F23H11.3"
/note="Hypothetical protein"
/codon_start=1
/protein_id="AAD39314.1"
/db_xref="PID:95080805"
/db_xref="GI:5080805"
/translation="MEDTKLGDKSSNVSLGSSCYOLDRYHATVLOTCSMDSOLKOL
HAFILRTTYPEEPATLPLYGKTILOLSSFSVDVNAFYRDSIENHSSPMWTLIRACA
HDSRKREAFMLYRKMLERGSSEPKHFFPYLKACAITFSSGKQVHCQIVHGGG
GDVYVNNGLIHLGSCGLDIARVFDMPERSLVSNWSMIDALVAFEGEYSALQLEFR
EMORSFEPDGYTGOSGLDLSGTNAHAFILKRCVDVAMADVLVNSLIEMY
CKGSLRMAEQVFOGMOKRDLASNMAMLIIGFATGRABEANFEDRMVKKRENVNPS
VTFGILACNHRGFVNGROYFDMMVADYCIPEPLIEHYGCIYDLIAAGYITTEADID
VMSMPKPDVATWRSILDACCKKASVLSSEIRANITGTEEDNESSKNGSCAYLL
SRVYASAKRNDVGIYRKLMSEHGIRKPGGCSLEINISHEFFRAGDTSHQTCQITO
QLKVIDLRLRIGYLPDRSQAPLVDAINDGSKSESLRLHSRLAIARGLINLPOTPI
RIFNLRLVNCNDCEHTKLISKVFNVEIIVDRVYFHHKDKSCSLDYM"
16582..18054
/gene="F23H11.4"
join(16582..1735,17613..18054)
/gene="F23H11.4"
/note="Very similar to hsp40, heat shock proteins."
/codon_start=1
/product="Putative heat shock protein"
/protein_id="AAD39315.1"

```

```

/db_xref="PID:95080806"
/db_xref="GI:5080806"
/translation="MGDYVYVNVNPSATEDLKKSYRLAMKMHDPDKNPTSJKOEA
EARKQISEADVSDPNKROIYQGEDGELATEAASQOHNYSGNNNYVAGRA
YYPDADIEDIEFFGASKEVDDGVGGGRKSAEASQNRKIPVRKAPALISKLA
CTLEELKKGGRKKKISRVPDGGKRPVEIKLIDITGKSGTITPERKNOBP
GVTADLITFIDENPHSYRKDNDLYDKKVSILELGTITSLTLDGNANITPVL
DIVPGEIVTSPSGMPTISKSGSRGLRINFELCFPSRLTSEQTKDLKRVLSGS"
complement(18603..19236)
/gene="F23H11.5"
complement(join(18603..18734,18816..18938,19102..19236))
/gene="F23H11.5"
/note="Similar to thioredoxins. Contains thioredoxin
family motif."
/codon_start=1
/product="Putative thioredoxin"
/protein_id="AAD39316.1"
/db_xref="PID:95080807"
/db_xref="GI:5080807"
/translation="MGSNVSVDVSHSMETTSNGEYVEIESRQMSLSDSKGSK
LVLTDFAVWCGPKAMEPRVREIASYSFAVFAVDYDRIMDVAGYRAITTPAFVF
VKRGEIDRVYGAKPDELVKIKEDHRV"
34973..38156
/gene="F23H11.6"
join(34973..35434,36790..36917,37322..38156)
/gene="F23H11.6"
/note="Similar to nitrate and oligopeptide transporters"
/codon_start=1
/protein_id="AAD39317.1"
/db_xref="PID:95080808"
/db_xref="GI:5080808"
/translation="MYCHVYSOGFTLLSYVAHLPLQKPPKCNPIDTCEAGKFKM
IFPMALVLVAGSCVPMNIAHGDADPSSHPKQSRKLSYFNAAFASWGLIL
TLVAVQVTHSGMDIGFVSAAMTMGILISVSGMYRNKRPRSITPLAHYVAI
LKRLASPSDPRMLHGDHVAANDVPSTLPHPRFPLDKACIKIDITAKESPMIL
CTVQVOVKTLISLVPISASTYFNILLOLOFESVOOQSNMTLSNFSFHPAL
OATPYMLIFLVPYDSFLVPAKRLGHSGSLPPLRIGLIGLFSFVSAMLEK
KRDSVLDGRILISFWITTPQFLRIGSEMTANGILFEFYKQSAKMESEFALITTC
SYSFGFSSVLVSVNKTSTSVDSGMGENDLNKRDLDLFYLLAVLSLNFESTY
LFWSRNWKISRRNNTVVDENT"
46258..50037
/gene="F23H11.7"
/note="Amp1."
join(46258..46294,46500..46615,46871..46966,47073..47129,
47228..47380,47466..47550,47704..47794,47873..48037,
48263..48382,48473..48548,48668..48839,48916..49385,
49532..49722,49869..50037)
/gene="F23H11.7"
/note="Rare gc intron splice site at 47794 is inferred
from alignment with cdna sequence."
/codon_start=1
/product="auxin response factor 1"
/protein_id="AAD39318.1"
/db_xref="PID:95080809"
/db_xref="GI:5080809"
/translation="MAASNSSGKPGCVLSDALCRELHMACAGPYLTLPREGRYVYF
PEGMEDLESMHGLEQMPSENLPSKILCKVINIRKREPEDEYTAOTITLPELD
QSELTDAVDEEKEKTVHSCFKTLTASDTSHGERSVARRADCLIFILDSQOP
MOELVAVLDLSENEHEHFRGQRLRLITGMSVFSKRLVAGDAFIFILGNEEL
RVGVRARROOTNIPSSVSISSHSHLVLTAHAALITGPIFSEYAPRISREFTVS
VNRLEKKTOLSLGMRFKRPGEEAPEKRGITTYGVQENISVHSDSEMSITVO
WDESSVFRPRERSPWLEPLVANSPTSSPOPORAKRRPRPGLSPATPGSPVTP
DGWKSSTADPSSVPLESPPAKATYFHGKNKFSVSGISGAFMPTNDAESFASFA
NNBESTEKQNGNVCARLFGELVENVWDCFSFASVSAVADQVPVPSFSGOOS
EPLINQSDIPSGSGDEKSLSPQSSQRCRSCCTKYMQSAYGARAIDLIRSECY
EDLEFKLEMEIDKIGELSESTKKQVYVYITDEDDMAVYGDPPNNEFCGAVRKIFITP
EYVKTLSPKKLAVNARMOLKADAEENGNEGSSMASGR"
complement(50415..55954)

```

Query Match 83.2%; Score 15.8; DB 8; Length 144422;
 Best Local Similarity 89.5%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```

Oy      1 AGGCTTCTTACGCATCA 19
Db      662 AGGCTTCTTACGCCTGA 644

RESULT 15
SPAC24H6/c
LOCUS   SPAC24H6      36833 bp      DNA      PLN      22-APR-1999
DEFINITION S.pombe chromosome I cosmid c24H6.
ACCESSION 254142
NID      9984697
VERSION  254142.1 GI:984697
KEYWORDS 40S ribosomal protein S9; 5S rRNA; cdc25; cullin 3; DAHP synthetase
          family; guanine nucleotide exchange factor; hexokinase; Homol D
          box; Homol E box; hsk1; M-phase inducer phosphatase; major
          facilitator family; phospho-3-deoxyheptonate aldolase;
          Rhodanese; rps9; ubiquitin activating enzyme; ubiquitin--protein
          ligase.
SOURCE   fission yeast.
ORGANISM Schizosaccharomyces pombe
          Eukaryota; Fungi; Ascomycota; Archiascomycetes;
          Schizosaccharomycetales; Schizosaccharomycetaceae;
          Schizosaccharomycetes.
REFERENCE 1 (bases 1 to 36833)
AUTHORS   Skelton,J. and Churcher,C.M.
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 36833)
AUTHORS   Barrell,B.G., Rajandream,M.A., Walsby,S.V. and Wood,V.
JOURNAL   Submitted (10-SEP-1995) Schizosaccharomyces pombe chromosome I
          Direct Submission
          sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
          CB10 1RQ E-mail: barrell@sanger.ac.uk
COMMENT  Details of yeast sequencing at the Sanger Centre are available on
          the World Wide Web.
          (URL: http://www.sanger.ac.uk/Projects/S_pombe/)
          Protein coding regions (CDS) have been predicted with the help of
          computer analysis using the Genetinder program in Pombase (an ACEDB
          database) with additional predictions for the branch-acceptor sites
          supplied by the program Sp3splice. CAUTION: It is possible that for
          any individual CDS we may have underestimated or overestimated the
          number of introns/exons or we may not have chosen the correct
          splice donor/acceptor sites.
          CDS are numbered using the following system eg SPAC5H10.01c. SP (S.
          pombe), A (Chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
          (complementary strand).
          The more significant matches with motifs in the PROSITE database
          are also included but some of these may be fortuitous.
          The length in codons is given for each CDS.
          IMPORTANT: This sequence MAY NOT be the entire insert of the
          sequenced clone. It may be shorter because we only sequence
          overlapping sections once, or longer, because we arrange for a
          small overlap between neighbouring submissions.
          Cosmid c24H6 is overlapped at the 5' end by cosmid c23E2.

FEATURES
SOURCE
Location/Qualifiers
1..36833
/organism="Schizosaccharomyces pombe"
/strain="972h-"
/db_xref="taxon:4896"
/chromosome="I"
/clone="cosmid c24H6"
/map="1L"
misc_feature
1
/note="nominal overlap with c23E2 EM:268887"
complement(30..800)
/gene="SPAC24H6.01c"
complement(join(30..478,554..800))
/gene="SPAC24H6.01c"
/note="SPAC24H6.01c, len:231aa; similarity: to YG1084C,
YG14 YEAST, HYPOTHETICAL PROTEIN, (560 aa);opt:300,
E():1.6e-15, (33.3% identity in 183 aa overlap)"
/codon_start=1
/product="hypothetical major facilitator family protein"

```

```

/misc_feature
complement(479..495)
/gene="SPAC24H6.01c"
/note="splice branch and acceptor sequence,
aactaactagcgtcag"
complement(348..553)
/gene="SPAC24H6.01c"
/note="splice donor sequence, gtaagt"
complement(1466..2037)
/gene="SPAC24H6.02c"
complement(join(1466..1742,1787..2037))
/gene="SPAC24H6.02c"
/note="SPAC24H6.02c, len:175aa"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAA90846.1"
/db_xref="PID:984698"
/db_xref="GI:984698"
/translation="MLRFREVDLESTKTERPNKSSRLSSNGSSHPSSSRLTV
RSAPVKSARGSTIEFYSVILITLACFKTHYVSSPKPKYKNEFKLKGWLEFG
QKYSADPQYSARENNPILLLIYIYNFMRKLVKLYFTNTNDELAIRNNYLCFSL
LFALVGTGVITVITLITNLISLKSINPLTLTWLIDISVVFKEFYAFKFS
SLHPGLGFLIRVY"
complement(1743..1759)
/gene="SPAC24H6.02c"
/note="splice branch and acceptor sequence,
tactaactaataagaag"
complement(1781..1786)
/gene="SPAC24H6.02c"
/note="splice donor sequence, gtaagt"
2201..4856
/gene="SPAC24H6.03"
join(2201..2242,2299..3683,3723..3832,3885..3949,
3994..4167,4219..4628,4685..4856)
/gene="SPAC24H6.03"
/note="SPAC24H6.03, len:785aa; similarity: to G3139079,
cullin 3, (768aa), fasta scores, opt:602, E():0, (37.3%
identity in 812 aa overlap)"
/codon_start=1
/product="cullin 3 homolog"
/protein_id="CAA90847.1"
/db_xref="PID:el312068"
/db_xref="PID:g3336937"
/db_xref="GI:3336937"
/translation="MORSAKLIRAPKFSANOVDPATHEVLORAIGDIFQKSTSL
SFEELYNATILVLHKGKLYNHVDVVISRLKEEVPATIKNYDSLGNALDLIR
KNDYSYTSWRSRLSAAHRLPSLSDNSKMSVLYKLYDKYSSAKVPNA
ENGIIYFREVVLNSFEIGECVETIILVYLEKGKTIKRLINDLIDMLNSPNS
KKEIYDVLAPKFLSTYRNFYEIESTVIGGVVYLKAEKRFEEKERSKNTYF
KTIASPLVSVEDELLSKHLDILENSTGFEFSDSNEFGIOLVYESPSRELVK
SKRYLAKYVAHGRKLINETTSQLEKMAVGRSSNATATLVQVAVLMDRLNTI
ISTTMDARSILNSDAVTFVQGYTRAEYISLFTDDNKKDKARAKISLEATQ
NSVTLFRFIEKDVFEKYRTHLAKRLNNRSISSDELMISRLKQDAVNTQKLE
GMFNDMLSOELLQYKNSALQSAKPALDNLNYSILASTFWPIDLSPKIKCNPKYL
LAQIDQFTDLYLSKRTGKRLMYSMGSADVRYVRKRYDLNVSTIASYLLLFQDL
KENOCILFEETLEKTNEVGDIRNLOSLACAKYKILLKDPKGRVAVAGKRYENEF
VSNARIKISTVATQTRVEDSEKRTLEKYEDESKRHQADCIYRVAKDRVYCEHNOIM
AEVTRQLPNRFHSPMKIKRRIETLIERETLQRODNGRITETLA"
2243..2248
/gene="SPAC24H6.03"
/note="splice donor sequence, gtaagt"
2286..2298
/gene="SPAC24H6.03"
/note="splice branch and acceptor sequence, tactaactag"

```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:35 ; Search time 255.05 Seconds

(without alignments)
18.638 Million cell updates/sec

Title: US-09-325-095-25

Sequence: 1 AGGCTTCTTTAGCCATCA 19

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	19	100.0	3412	1 051156	Human p110 CDNA. R
C 2	19	100.0	3207	1 051155	p110 CDNA. Recombi
C 3	19	100.0	3428	1 057012	Pdins 3-kinase 11
C 4	15.4	81.1	2880	1 X14318	H. pylori GHPO 127
C 5	15	78.9	8196	1 V09699	Porcine retrovirus
C 6	15	78.9	1974	1 V09703	Porcine retrovirus
C 7	15	78.9	8209	1 V09700	Porcine retrovirus
C 8	14.8	77.9	1325	1 036666	Potato tuber PPO C
C 9	14.8	77.9	1670	1 056132	Potato polyphenol
C 10	14.8	77.9	682	1 T78398	Potato polyphenol
C 11	14.8	77.9	2370	1 V48311	Human cell-volume
C 12	14.8	77.9	2346	1 V74190	Human cell-volume
C 13	14.8	77.9	3166	1 X14374	Human cell-volume
C 14	14.8	77.9	2085	1 X01362	Nucleoside permeas
C 15	14.8	77.9	1907	1 X20307	Borrelia burgdorferi
C 16	14.8	77.9	1924	1 V72243	G. max SBP2 CDNA.
C 17	14.4	75.8	17041	1 021065	Genomic DNA of hum
C 18	14.4	75.8	3980	1 062696	Sequence of human
C 19	14.4	75.8	17041	1 067057	PACBP38 DNA. PACA
C 20	14.4	75.8	6000	1 T06308	Arabidopsis pathog
C 21	14.4	75.8	7000	1 T37476	RPP5 downy mildew
C 22	14.4	75.8	2329	1 X14596	H. pylori GHPO 175
C 23	14.2	74.7	1757	1 049428	Cytochrome P450 ho
C 24	14.2	74.7	2652	1 051233	Sequence encoding
C 25	14.2	74.7	4008	1 064827	Leukotoxin Apptin
C 26	14.2	74.7	110000	1 T58840.4	Continuation (5 of
C 27	14.2	74.7	7721	1 T73320	ApptinABD gene. P
C 28	14.2	74.7	1757	1 T94666	Petunia flavonoid
C 29	14.2	74.7	2448	1 V01882	Human nel-related
C 30	14.2	74.7	3198	1 V01883	Human nel-related
C 31	14.2	74.7	409	1 V75559	Staphylococcus aur
C 32	14.2	74.7	1903	1 X14490	H. pylori GHPO 113
C 33	14.2	74.7	1227	1 X14478	H. pylori GHPO 108
C 34	14.2	74.7	1465	1 X14121	H. pylori GHPO 153
C 35	14.2	74.7	1262	1 X27327	Human secreted pro
C 36	14.2	74.7	8241	1 N81439	Factor VIII CDNA. 1
C 37	14.2	74.7	8241	1 N81096	CDNA sequence enco
C 38	14.2	74.7	7059	1 N81543	Human factor VIII-
C 39	14.2	74.7	7440	1 N80689	Sequence encoding
C 40	14.2	74.7	7587	1 Q15133	Zuccchini ACC synth
C 41	14.2	74.7	15155	1 N50107	DNA encoding Facto
C 42	14.2	74.7	3852	1 N50094	DNA encoding Facto
C 43	14.2	74.7	8975	1 N50054	Human factor VIII

ALIGNMENTS

C 44 14 73.7 7440 1 N50375 DNA sequence encod
C 45 14 73.7 4464 1 X22901 MO 99/07855 SeqID

```

RESULT 1
ID 051156/c
AC 051156:
DE 12-APR-1994 (first entry)
KW Human p110 CDNA.
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ds.
OS Human.
FH Key
FT cas
FT /tag- a
FT /note- "PI3- kinase p110"

MO321328-A.
PD 28-OCT-1993.
PF 13-APR-1993; G00761.
PR 13-APR-1992; GB-008135.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI Parker PJ, Volinia S, Waterfield MD;
DR WPI: 93-351738/44.
DR P-PDB: R43342.
PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PT activity; useful for controlling cell proliferation
PS Claim 7; Fig 16; 146bp; English.
CC Southern blot analysis was performed using a bovine CDNA probe contg
CC a fragment of a PI3-kinase-encoding sequence and human CDNA isolated
CC from a CDNA library constructed from mRNA isolated from the human
CC cell line K562. Positive clones were sequenced to give the human
CC PI3 kinase p110 sequence shown. This sequence has 95 percent
CC homology with the bovine sequence. The domain encoding residues 19-
CC 100 of human p110 is sufficient to encode the kinase which will
CC associate with the p85 kinase subunit. The gene may be used to
CC provide a protein with PI3 kinase activity, and is useful for
CC screening for (ant)agonists of PI3 kinase activity which could be
CC useful for stimulation or inhibition of cell proliferation and hence
CC prophylaxis or therapy. Platelet or neutrophil activity or blood
CC glucose levels can be controlled using the kinase.
CC See also 051155 and Q57522-3.
SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 100.0%; Score 19; DB 1; Length 3412;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
   |||||||||||||||
DB 878 AGGCTTCTTTAGCCATCA 860

RESULT 2
ID 051155/c
AC 051155:
DE 12-APR-1994 (first entry)
KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW ss.
FH Key
FT cas
FT /tag- a
FT /note- a

```

```

FM      /note- "P13- kinase p110"
PM      WO9321328-A.
PD      28-OCT-1993.
PF      13-APR-1993; G00761.
PR      13-APR-1992; GB-008135.
PA      (LUDWIG) LUDWIG INST CANCER RES.
PI      Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI      Parke PJ, Volinia S, Waterfield MD;
PI      WPI: 93-351738/44.
DR      P-PSDB; R43341.
PT      Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PS      activity, useful for controlling cell proliferation
PS      Claim 7; Fig 9; 146pp; English.
CC      Southern blot analysis was performed using a bovine cDNA probe contg.
CC      a fragment of a P13-kinase-encoding sequence and human cDNA isolated
CC      from a cDNA library constructed from mRNA isolated from the human
CC      cell line KG1a. Positive clones were sequenced to give the human
CC      P13 kinase p110 sequence shown. This sequence has 95 percent
CC      homology with the bovine sequence. The domain encoding residues 19-
CC      100 of human p110 is sufficient to encode the kinase which will
CC      associate with the p85 kinase subunit. The gene may be used to
CC      provide a protein with P13 kinase activity, and is useful for
CC      screening for (ant)agonists of P13 kinase activity which could be
CC      useful for stimulation or inhibition of cell proliferation and hence
CC      prophylaxis or therapy. Platelet or neutrophil activity or blood
CC      glucose levels can be controlled using the kinase.
CC      See also Q51156, Q59012-23 and Q57522-3.
SQ      Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match          100.0%; Score 19; DB 1; Length 3207;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCGTTCTTACCATCA 19
        |||||||
Db      878 AGCGTTCTTACCATCA 860

RESULT 3
O57012/c
ID      O57012 standard; cDNA to mRNA; 3498 BP.
AC      O57012:
DE      31-AUG-1994 (first entry)
DE      Ptdins 3-kinase 110 kD catalytic subunit cDNA.
KW      110 kD catalytic subunit; phosphatidylinositol 3-kinase;
KW      transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW      Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
KW      blood vessel plaques; ss.
OS      Bos taurus.
FH      Key
FT      cds
FT      Location/Qualifiers
FT      1..3207
FT      /*tag= a
FT      /*product= p110

WO9403409-A.
PD      17-FEB-1994.
PF      05-AUG-1993; G01651.
PR      05-AUG-1992; GB-016654.
PA      (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI      Goode NT, Nurse PM, Parker PJ, Waterfield MD;
PI      WPI: 94-065697/08.
DR      P-PSDB; R46294.
PT      Eukaryotic cells transformed with mammalian phospholipid or
PT      protein kinase DNA - useful in assays for compounds involved in
PT      cell growth regulation and for treating cancers
PS      Disclosure: Fig 1; 71pp; English.
CC      This sequence encodes the 110 kD catalytic subunit of the phosphatidylinositol
CC      (Ptdins) 3-kinase. This sequence was transformed into
CC      Schizosaccharomyces pombe cells under the regulatory control of the
CC      nmt promoter in an embodiment of the invention. In the presence of
CC      thiamine the promoter is inactive and the cells carrying the Ptdins
CC      catalytic subunit plasmid grow as the parental strain. In the absence
CC      of thiamine the nmt promoter functions and the Ptdins 3-kinase

```

```

CC      catalytic subunit in induced. Ptdins activity is substantially
CC      increased under these conditions. Cells containing constructs such as
CC      this, are useful in assays for detecting compounds involved in cell
CC      growth regulation. It is also used as the basis for detecting
CC      compounds for treating cancers and the formation of blood vessel
CC      plaques.
SQ      Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match          100.0%; Score 19; DB 1; Length 3498;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGCGTTCTTACCATCA 19
        |||||||
Db      878 AGCGTTCTTACCATCA 860

RESULT 4
X14318/c
ID      X14318 standard; DNA; 2880 BP.
AC      X14318;
DE      31-MAR-1999 (first entry)
DE      H. pylori GHPD 127 gene.
KW      GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW      peptic ulcer disease; ss.
OS      Helicobacter pylori.
FH      Key
FT      CDS
FT      Location/Qualifiers
FT      16..2817
FT      /*tag= a

WO9843478-A1.
PD      08-OCT-1998.
PF      01-APR-1998; U06371.
PR      29-JUL-1997; US-902615.
PR      01-APR-1997; US-833457.
PR      24-JUN-1997; US-881227.
PA      (HUMA-) HUMAN GENOME SCI INC.
PA      (IMMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI      Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
PI      WPI: 98-542293/46.
DR      P-PSDB; W98599.
PT      New isolated Helicobacter polynucleotides - used to develop products
PT      for the diagnosis, prevention and treatment of Helicobacter
PT      infections and gastrointestinal diseases
PS      Claim 1; Page 1195-1199; 2054pp; English.
CC      This sequence represents a polynucleotide of the invention. It was
CC      isolated from Helicobacter pylori and encodes a H.pylori GHPD protein.
CC      The polypeptides can be used for preventing or treating Helicobacter
CC      infections, and gastroduodenal diseases associated with these
CC      infections, including acute, chronic, and atrophic gastritis, and peptic
CC      ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC      for the production of antibodies. The products can also be used for
CC      detection and diagnosis.
SQ      Sequence 2880 BP; 997 A; 553 C; 589 G; 741 T;

Query Match          81.1%; Score 15.4; DB 1; Length 2880;
Best Local Similarity 94.1%; Pred. No. 36;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 AGCGTTCTTACCAT 17
        |||||||
Db      2690 AGCGTTCTTATCCAT 2674

RESULT 5
WO9699/c
ID      V09699 standard; DNA; 8196 BP.
AC      V09699;
DE      19-MAY-1998 (first entry)
DE      Porcine retrovirus DNA encoding, GAG, POL and ENV.
KW      Porcine retrovirus; POLV; POL protein; ENV protein; GAG protein;
KW      vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds...

```

```

OS Porcine retrovirus.
FH Key Location/Qualifiers
FT CDS 576..2126
      /tag=a
      /product=GAG protein
      /note="viral core polypeptide"
FT FT 2143..5733
      /tag=b
      /product=POL protein
      /note="polymerase peptide sequence as given in
      specification"
FT FT 5606..7576
      /tag=c
      /product=ENV protein
      /note="envelope protein"
FT FT
FT FT
FT FT
PD WO9740167-A1.
PD 30-OCT-1997.
PF 18-APR-1997; G01087.
PR 10-FEB-1997; GB-002668.
PR 19-APR-1996; GB-008164.
PA (IMUT-) IMUTRAN LTD.
PA (OONE-) O-ONE BIOTECH LTD.
PI Galbraith DN, Haworth C, Lees GM, Smith KT;
DR WPI: 97-535851/49.
PT Polynucleotide encoding porcine retrovirus expression product -
PT useful to develop products for use in vaccines, diagnosis and
PT xeno-transplantation
PS Claim 4; Fig 2; 69pp; English.
CC This DNA sequence encodes the porcine retrovirus (POEV) viral core
CC polypeptide (GAG), polymerase (POL) and envelope (ENV) proteins. These
CC proteins can be used to develop viral vaccines, antisense nucleic acids,
CC ribozymes and other antiviral agents. They can also be used in
CC xeno-transplantation technology and as diagnostic tools.
SQ Sequence 8196 BP; 2165 A; 2061 C; 2147 G; 1820 T;

Query Match 78.9%; Score 15; DB 1; Length 8196;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CTTCTTAGCCATC 18
   |||||
Db 6624 CTTCTTAGCCATC 6610

RESULT 6
V03703/c
ID V03703 standard; DNA: 1974 BP.
AC V03703;
DT 20-MAY-1998 (first entry)
DE Porcine retrovirus Raj1 clone ENV DNA.
DE Porcine retrovirus; POEV; ENV protein; envelope protein; vaccine;
KW diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
US Porcine retrovirus.
FH Key Location/Qualifiers
FT CDS 1..1974
      /tag=a
      /product=ENV protein
      /note="envelope protein"
FT FT
FT FT
FT FT
PD WO9740167-A1.
PD 30-OCT-1997.
PF 18-APR-1997; G01087.
PR 10-FEB-1997; GB-002668.
PR 19-APR-1996; GB-008164.
PA (IMUT-) IMUTRAN LTD.
PA (OONE-) O-ONE BIOTECH LTD.
PI Galbraith DN, Haworth C, Lees GM, Smith KT;
DR WPI: 97-535851/49.
PT Polynucleotide encoding porcine retrovirus expression product -
PT useful to develop products for use in vaccines, diagnosis and
PT xeno-transplantation
PS; Claim 6; Fig 4; 69pp; English.

```

```

CC This sequence encodes the porcine retrovirus envelope (ENV) protein
CC isolated from the human cell line Raj1. Such viral proteins can be used
CC to develop viral vaccines, antisense nucleic acids, ribozymes and other
CC antiviral agents. They can also be used in xeno-transplantation
CC technology and as diagnostic tools.
SQ Sequence 1974 BP; 534 A; 489 C; 483 G; 468 T;

Query Match 78.9%; Score 15; DB 1; Length 1974;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CTTCTTAGCCATC 18
   |||||
Db 1022 CTTCTTAGCCATC 1008

RESULT 7
V03700/c
ID V03700 standard; DNA: 8209 BP.
AC V03700;
DT 20-MAY-1998 (first entry)
DE Porcine retrovirus DNA encoding the LTR and GAG, POL and ENV proteins.
DE Porcine retrovirus; POEV; POL protein; ENV protein; GAG protein;
KW viral core polypeptide; polymerase protein; envelope protein;
KW vaccine; diagnosis; xenotransplantation; prophylactic; therapeutic; ds.
OS Porcine retrovirus.
FH Key Location/Qualifiers
FT misc-feature 1..61
      /tag=a
      /note="R LTR domain"
FT LTR 1..588
      /tag=b
      /product=GAG protein
      /note="viral core polypeptide"
FT misc-feature 62..143
      /tag=c
      /note="U5 LTR domain"
FT CDS 588..2162
      /tag=d
      /product=ENV protein
      /note="polymerase peptide"
FT CDS 2163..5747
      /tag=e
      /product=POL protein
      /note="envelope protein"
FT CDS 5620..7590
      /tag=f
      /product=ENV protein
      /note="envelope protein"
FT LTR 7591..8209
      /tag=g
      /tag=h
      /tag=i
      /tag=j
      /tag=k
      /tag=l
      /tag=m
      /tag=n
      /tag=o
      /tag=p
      /tag=q
      /tag=r
      /tag=s
      /tag=t
      /tag=u
      /tag=v
      /tag=w
      /tag=x
      /tag=y
      /tag=z
      /tag=aa
      /tag=ab
      /tag=ac
      /tag=ad
      /tag=ae
      /tag=af
      /tag=ag
      /tag=ah
      /tag=ai
      /tag=aj
      /tag=ak
      /tag=al
      /tag=am
      /tag=an
      /tag=ao
      /tag=ap
      /tag=aq
      /tag=ar
      /tag=as
      /tag=at
      /tag=au
      /tag=av
      /tag=aw
      /tag=ax
      /tag=ay
      /tag=az
      /tag=ba
      /tag=bb
      /tag=bc
      /tag=bd
      /tag=be
      /tag=bf
      /tag=bg
      /tag=bh
      /tag=bi
      /tag=bj
      /tag=bk
      /tag=bl
      /tag=bm
      /tag=bn
      /tag=bo
      /tag=bp
      /tag= bq
      /tag=br
      /tag=bs
      /tag=bt
      /tag=bu
      /tag=bv
      /tag=bw
      /tag=bx
      /tag=by
      /tag=bz
      /tag=ca
      /tag=cb
      /tag=cc
      /tag=cd
      /tag=ce
      /tag=cf
      /tag=cg
      /tag=ch
      /tag=ci
      /tag=cj
      /tag=ck
      /tag=cl
      /tag=cm
      /tag=cn
      /tag=co
      /tag=cp
      /tag=cq
      /tag=cr
      /tag=cs
      /tag=ct
      /tag=cu
      /tag=cv
      /tag=cw
      /tag=cx
      /tag=cy
      /tag=cz
      /tag=da
      /tag=db
      /tag=dc
      /tag=dd
      /tag=de
      /tag=df
      /tag=dg
      /tag=dh
      /tag=di
      /tag=dj
      /tag=dk
      /tag=dl
      /tag=dm
      /tag=dn
      /tag=do
      /tag=dp
      /tag=dq
      /tag=dr
      /tag=ds
      /tag=dt
      /tag=du
      /tag=dv
      /tag=dw
      /tag=dx
      /tag=dy
      /tag=dz
      /tag=ea
      /tag=eb
      /tag=ec
      /tag=ed
      /tag=ee
      /tag=ef
      /tag=eg
      /tag=eh
      /tag=ei
      /tag=ej
      /tag=ek
      /tag=el
      /tag=em
      /tag=en
      /tag=eo
      /tag=ep
      /tag=eq
      /tag=er
      /tag=es
      /tag=et
      /tag=eu
      /tag=ev
      /tag=ew
      /tag=ex
      /tag=ey
      /tag=ez
      /tag=fa
      /tag=fb
      /tag=fc
      /tag=fd
      /tag=fe
      /tag=ff
      /tag=fg
      /tag=fh
      /tag=fi
      /tag=fj
      /tag=fk
      /tag=fl
      /tag=fm
      /tag=fn
      /tag=fo
      /tag=fp
      /tag=fq
      /tag=fr
      /tag=fs
      /tag=ft
      /tag=fu
      /tag=fv
      /tag=fw
      /tag=fx
      /tag=fy
      /tag=fz
      /tag=ga
      /tag=gb
      /tag=gc
      /tag=gd
      /tag=ge
      /tag=gf
      /tag=gg
      /tag=gh
      /tag=gi
      /tag=gj
      /tag=gk
      /tag=gl
      /tag=gm
      /tag=gn
      /tag=go
      /tag=gp
      /tag=gq
      /tag=gr
      /tag=gs
      /tag=gt
      /tag=gu
      /tag=gw
      /tag=gx
      /tag=gy
      /tag=gz
      /tag=ha
      /tag=hb
      /tag=hc
      /tag=hd
      /tag=he
      /tag=hf
      /tag=hg
      /tag=hh
      /tag=hi
      /tag=hj
      /tag=hk
      /tag=hl
      /tag=hm
      /tag=hn
      /tag=ho
      /tag=hp
      /tag=hq
      /tag=hr
      /tag=hs
      /tag=ht
      /tag=hu
      /tag=hw
      /tag=hx
      /tag=hy
      /tag=hz
      /tag=ia
      /tag=ib
      /tag=ic
      /tag=id
      /tag=ie
      /tag=if
      /tag=ig
      /tag=ih
      /tag=ii
      /tag=ij
      /tag=ik
      /tag=il
      /tag=im
      /tag=in
      /tag=io
      /tag=ip
      /tag=iq
      /tag=ir
      /tag=is
      /tag=it
      /tag=iu
      /tag=iv
      /tag=iw
      /tag=ix
      /tag=iy
      /tag=iz
      /tag=ja
      /tag=jb
      /tag=jc
      /tag=jd
      /tag=je
      /tag=jf
      /tag=jg
      /tag=jh
      /tag=ji
      /tag=jj
      /tag=jk
      /tag=jl
      /tag=jm
      /tag=jn
      /tag=jo
      /tag=jp
      /tag=jq
      /tag=jr
      /tag=js
      /tag=jt
      /tag=ju
      /tag=jw
      /tag=jx
      /tag= jy
      /tag=jz
      /tag=ka
      /tag=kb
      /tag=kc
      /tag=kd
      /tag=ke
      /tag=kf
      /tag=kg
      /tag=kh
      /tag=ki
      /tag=kj
      /tag=kk
      /tag=kl
      /tag=km
      /tag=kn
      /tag=ko
      /tag=kp
      /tag=kq
      /tag=kr
      /tag=ks
      /tag=kt
      /tag=ku
      /tag=kw
      /tag=kx
      /tag=ky
      /tag=kz
      /tag=la
      /tag=lb
      /tag=lc
      /tag=ld
      /tag=le
      /tag=lf
      /tag=lg
      /tag=lh
      /tag=li
      /tag=lj
      /tag=lk
      /tag=ll
      /tag=lm
      /tag=ln
      /tag=lo
      /tag=lp
      /tag=lq
      /tag=lr
      /tag=ls
      /tag=lt
      /tag=lu
      /tag=lw
      /tag=lx
      /tag=ly
      /tag=lz
      /tag=ma
      /tag=mb
      /tag=mc
      /tag=md
      /tag=me
      /tag=mf
      /tag=mg
      /tag=mh
      /tag=mi
      /tag=mj
      /tag=mk
      /tag=ml
      /tag=mm
      /tag=mn
      /tag=mo
      /tag=mp
      /tag=mq
      /tag=mr
      /tag=ms
      /tag=mt
      /tag=mu
      /tag=mw
      /tag=mx
      /tag=my
      /tag=mz
      /tag=na
      /tag=nb
      /tag=nc
      /tag=nd
      /tag=ne
      /tag=nf
      /tag=ng
      /tag=nh
      /tag=ni
      /tag=nj
      /tag=nk
      /tag=nl
      /tag=nm
      /tag=nn
      /tag=no
      /tag=np
      /tag=nq
      /tag=nr
      /tag=ns
      /tag=nt
      /tag=nu
      /tag=nw
      /tag=nx
      /tag=ny
      /tag=nz
      /tag=oa
      /tag=ob
      /tag=oc
      /tag=od
      /tag=oe
      /tag=of
      /tag=og
      /tag=oh
      /tag=oi
      /tag=oj
      /tag=ok
      /tag=ol
      /tag=om
      /tag=on
      /tag=oo
      /tag=op
      /tag=oq
      /tag=or
      /tag=os
      /tag=ot
      /tag=ou
      /tag=ow
      /tag=ox
      /tag=oy
      /tag=oz
      /tag=pa
      /tag=pb
      /tag=pc
      /tag=pd
      /tag=pe
      /tag=pf
      /tag=pg
      /tag=ph
      /tag=pi
      /tag=pj
      /tag=pk
      /tag=pl
      /tag=pm
      /tag=pn
      /tag=po
      /tag=pp
      /tag=pq
      /tag=pr
      /tag=ps
      /tag=pt
      /tag=pu
      /tag=pw
      /tag=px
      /tag=py
      /tag=pz
      /tag=qa
      /tag=qb
      /tag=qc
      /tag=qd
      /tag=qe
      /tag=qf
      /tag=qg
      /tag=qh
      /tag=qi
      /tag=qj
      /tag=qk
      /tag=ql
      /tag=qm
      /tag=qn
      /tag=qo
      /tag=qp
      /tag=qq
      /tag=qr
      /tag=qs
      /tag=qt
      /tag=qu
      /tag=qw
      /tag=qx
      /tag=qy
      /tag=qz
      /tag=ra
      /tag=rb
      /tag=rc
      /tag=rd
      /tag=re
      /tag=rf
      /tag=rg
      /tag=rh
      /tag=ri
      /tag=rj
      /tag=rk
      /tag=rl
      /tag=rm
      /tag=rn
      /tag=ro
      /tag=rp
      /tag=rq
      /tag=rr
      /tag=rs
      /tag=rt
      /tag=ru
      /tag=rw
      /tag=rx
      /tag=ry
      /tag=rz
      /tag=sa
      /tag=sb
      /tag=sc
      /tag=sd
      /tag=se
      /tag=sf
      /tag=sg
      /tag=sh
      /tag=si
      /tag=sj
      /tag=sk
      /tag=sl
      /tag=sm
      /tag=sn
      /tag=so
      /tag=sp
      /tag=sq
      /tag=sr
      /tag=ss
      /tag=st
      /tag=su
      /tag=sw
      /tag=sx
      /tag=sy
      /tag=sz
      /tag=ta
      /tag=tb
      /tag=tc
      /tag=td
      /tag=te
      /tag=tf
      /tag=tg
      /tag=th
      /tag=ti
      /tag=tj
      /tag=tk
      /tag=tl
      /tag=tm
      /tag=tn
      /tag=to
      /tag=tp
      /tag=tq
      /tag=tr
      /tag=ts
      /tag=tt
      /tag=tu
      /tag=tw
      /tag=tx
      /tag=ty
      /tag=tz
      /tag=ua
      /tag=ub
      /tag=uc
      /tag=ud
      /tag=ue
      /tag=uf
      /tag=ug
      /tag=uh
      /tag=ui
      /tag=uj
      /tag=uk
      /tag=ul
      /tag=um
      /tag=un
      /tag=uo
      /tag=up
      /tag=uq
      /tag=ur
      /tag=us
      /tag=ut
      /tag=uu
      /tag=uw
      /tag=ux
      /tag=uy
      /tag=uz
      /tag=va
      /tag=vb
      /tag=vc
      /tag=vd
      /tag=ve
      /tag=vf
      /tag=vg
      /tag=vh
      /tag=vi
      /tag=vj
      /tag=vk
      /tag=vl
      /tag=vm
      /tag=vn
      /tag=vo
      /tag=vp
      /tag=vq
      /tag=vr
      /tag=vs
      /tag=vt
      /tag=vu
      /tag=vw
      /tag=vx
      /tag=vy
      /tag=vz
      /tag=wa
      /tag=wb
      /tag=wc
      /tag=wd
      /tag=we
      /tag=wf
      /tag=wg
      /tag=wh
      /tag=wi
      /tag=wj
      /tag=wk
      /tag=wl
      /tag=wm
      /tag=wn
      /tag=wo
      /tag=wp
      /tag=wq
      /tag=wr
      /tag=ws
      /tag=wt
      /tag=wu
      /tag=ww
      /tag=wx
      /tag=wy
      /tag=wz
      /tag=xa
      /tag=xb
      /tag=xc
      /tag=xd
      /tag=xe
      /tag=xf
      /tag=xg
      /tag=xh
      /tag=xi
      /tag=xj
      /tag=xk
      /tag=xl
      /tag=xm
      /tag=xn
      /tag=xo
      /tag=xp
      /tag=xq
      /tag=xr
      /tag=xs
      /tag=xt
      /tag=xu
      /tag=xw
      /tag=xx
      /tag=xy
      /tag=xz
      /tag=ya
      /tag=yb
      /tag=yc
      /tag=yd
      /tag=ye
      /tag=yf
      /tag=yg
      /tag=yh
      /tag=yi
      /tag=yj
      /tag=yk
      /tag=yl
      /tag=ym
      /tag=yn
      /tag=yo
      /tag=yp
      /tag=yq
      /tag=yr
      /tag=ys
      /tag=yt
      /tag=yu
      /tag=yw
      /tag=yy
      /tag=yz
      /tag=za
      /tag=zb
      /tag=zc
      /tag=zd
      /tag=ze
      /tag=zf
      /tag=zg
      /tag=zh
      /tag=zi
      /tag=zj
      /tag=zk
      /tag=zl
      /tag=zm
      /tag=zn
      /tag=zo
      /tag=zp
      /tag=zq
      /tag=zr
      /tag=zs
      /tag=zt
      /tag=zu
      /tag=zw
      /tag=zx
      /tag=zy
      /tag=zz

```



```

FT misc_feature 1173
FT /*tag- b
FT /label- site_E
FT /note- "insertion site"
FT misc_feature 1645
FT /*tag- c
FT /label- site_F
FT /note- "insertion site"
PN WO9403607-A.
PD 17-FEB-1994.
PF 27-JUL-1993; E01988.
PR 30-JUL-1992; EP-402199.
PA (KEYG-) KEYGENE NV.
PI Bachem C, Zabeau M,
DR WPI: 94-065696/08.
PT DNA constructs conty. antisense polyphenol oxidase genes - used
PT for producing plants, partic. potatoes with reduced
PT susceptibility to bruising and browning
PS Claim 11; Fig 1b; 30pp; English.
CC A potato tuber cDNA clone bank was made and screened with a probe
CC corresp. to the sequence for a leaf PPO gene. The isolated clones
CC clones from each class were PKG45-8 (Q56131) and PKG59-4 (Q56132).
CC The latter is thought to be uniquely expressed in tubers. Antisense
CC constructs were made using the two full-length clones. The
CC constructs were introduced into 2 different lines of potato via
CC Agrobacterium tumefaciens; in field trials, potatoes derived from
CC regenerants showed reduced bruising.
SQ Sequence 1670 BP; 480 A; 337 C; 373 G; 480 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1670;
Best Local Similarity 88.9%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCGTTCTTTAGCCATCA 19
DB 1058 GGTTCCTTTGCGCATCA 1075

RESULT 10
T78398
ID T78398 standard; DNA; 682 BP.
AC T78398;
DT 23-MAR-1998 (first entry)
DE Potato polyphenol oxidase GPO6 genomic DNA.
KW polyphenol oxidase; PPO; browning; fruit; vegetable;
KW genomic DNA amplification; ds.
OS Solanum tuberosum.
PN WO9729193-A1.
PD 14-AUG-1997.
PF 24-JAN-1997; AU0041.
PR 16-SEP-1996; AU-002361.
PR 05-FEB-1996; AU-007856.
PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI Robinson SP;
DR WPI: 97-415348/38.
DR P-PSDB: W23687.
PT Preparing nucleic acid encoding polyphenol oxidase - by genomic DNA
PT amplification, useful to control browning reactions in fruit and
PT vegetables
PS Claim 40; Fig 24; 53pp; English.
CC A method has been developed for preparing a nucleic acid sequence
CC encoding polyphenol oxidase (PPO), or a fragment or derivative. The
CC method comprises amplifying genomic DNA isolated from plant tissue with
CC sense and antisense primers corresponding to conserved PPO gene regions.
CC The present sequence encodes a specifically claimed polyphenol oxidase.
CC Sense nucleic acid sequences can be used to increase or, by
CC co-suppression, decrease PPO activity in plants, while antisense
CC nucleic acid sequences reduce activity. Control of PPO activity allows
CC browning reactions in fruit and vegetables to be controlled, while
CC avoiding the need for chemicals, e.g. sulphur dioxide. Many PPO genes
CC lack introns, and can therefore be amplified directly from genomic DNA.

```

```

CC eliminating the need for separation of RNA and synthesis of cDNA. Also,
CC only small samples are needed and fragment size can be predicted,
CC allowing bands of appropriate size to be selected for cloning.
SQ Sequence 682 BP; 186 A; 138 C; 158 G; 200 T;

Query Match 77.9%; Score 14.8; DB 1; Length 682;
Best Local Similarity 88.9%; Pred. No. 64;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GCGTTCTTTAGCCATCA 19
DB 557 GGTTCCTTTGCGCATCA 574

RESULT 11
V48311/c
ID V48311 standard; cDNA; 2370 BP.
AC V48311;
DT 16-NOV-1998 (first entry)
DE Human cell-volume regulating kinase h-sgk.
KW ss: human; cell-volume; Kinase; h-sgk; diabetes mellitus;
KW renal insufficiency; inflammation; Alzheimer's disease.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 43..138
FT /*tag- a
FT /product= "Kinase h-sgk"
PN EP-861896-A2.
PD 02-SEP-1998.
PF 27-JAN-1998; 101338.
PR 28-FEB-1997; DE-008173.
PA (DADE-) DADE BEHRING MARBURG GMBH.
PI Lang F, Waldegger S,
DR WPI: 98-449109/39.
DR P-PSDB: W77219.
DE New nucleic acid encoding cell-volume regulating kinase h-sgk and
DE related proteins - used for diagnosis and treatment of diseases
PT involving changes in cell volume, e.g. renal insufficiency,
PT inflammation, infections etc.
PS Disclosure: Fig 1; 15pp; German.
CC The human cell-volume regulating kinase h-sgk is inhibited by the
CC swelling of cells (or presence of urea), whereas cell shrinkage
CC stimulates its expression. The nucleic acid h-sgk, and fragments, are
CC particularly used to detect changes in cell volume, specifically for
CC diagnosis of conditions that involve such changes, e.g. hyper- and hypo-
CC natremia, diabetes mellitus, renal insufficiency, hypercatabolism,
CC hepatic encephalopathy, inflammation, microbial/viral infection, fructose
CC intolerance, hyper- and hypo-glycaemia and Alzheimer's disease.
CC The nucleic acid, protein and products including receptors that bind
CC h-sgk, can be used to treat these disorders.
SQ Sequence 2370 BP; 636 A; 517 C; 513 G; 704 T;

Query Match 77.9%; Score 14.8; DB 1; Length 2370;
Best Local Similarity 88.9%; Pred. No. 72;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGCGTTCTTTAGCCATC 18
DB 1994 AGCGTTCTTTAGCCATC 1977

RESULT 12
V74190/c
ID V74190 standard; DNA; 2346 BP.
AC V74190;
DT 15-MAR-1999 (first entry)
DE Human sgk DNA.
KW Serum glucocorticoid regulated kinase; sgk; human; treatment; inhibitor;
KW serine/threonine protein kinase family; antagonist; diabetic nephropathy;
KW chronic renal failure; inflammation; Alzheimer's disease; wound; ss.
OS Homo sapiens.

```

FH Key Location/Qualifiers
 FT CDS 45..1340
 FT /*tag= a
 FT /product= "sgk"
 FT /transl_except= (pos:1185..1187, aa:asp)
 PN EP-887081-A2.
 PD 30-DEC-1998.
 PR 27-MAY-1998; 304189.
 PR 27-JUN-1997; US-051124.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Kumar JM;
 DR P-PSDB; W90139.
 PT Treating chronic renal failure, diabetic nephropathy and Alzheimer's
 PT disease - by administration of nucleic acids and antagonists which
 PT inhibit activity or expression of human serum glucocorticoid
 PT regulated kinase (sgk), a serine/threonine protein kinase
 PS Disclosure: Page 14-15; 17pp; English.
 CC This sequence encodes a novel human serum glucocorticoid regulated kinase
 CC (sgk) protein which is a member of the serine/threonine protein kinase
 CC family. This protein is used for the treatment of a subject having need
 CC to inhibit/antagonise activity or expression of human sgk polypeptide
 CC e.g. for the treatment of chronic renal failure, diabetic nephropathy,
 CC inflammation, Alzheimer's disease and wounds.
 SO Sequence 2346 BP; 608 A; 520 C; 518 G; 700 T;

Query Match 77.9%; Score 14.8; DB 1; Length 2346;
 Best Local Similarity 88.9%; Pred. No. 72;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTGGCATTG 18
 Db 1999 AGGCTTCTTGGCATTG 1982

RESULT 13
 ID X14274
 AC X14274 standard; DNA; 3166 BP.
 DT 31-MAR-1999 (first entry)
 DE H. pylori GHP0 1661 gene.
 KW GHP0 protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease; ss.
 OS Helicobacter pylori.
 FH Key Location/Qualifiers
 FT CDS 51..3116
 FT /*tag= a
 PN M09843478-A1.
 PD 08-OCT-1998.
 PR 01-APR-1998; U06371.
 PR 29-JUL-1997; US-902615.
 PR 01-APR-1997; US-833457.
 PR 24-JUN-1997; US-881227.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (IMMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 PI Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
 DR WPI; 98-342293/46.
 DR P-PSDB; W98555.
 PT New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 PS Claim 1; Page 1082-1087; 2054pp; English.
 CC This sequence represents a polynucleotide of the invention. It was
 CC isolated from Helicobacter pylori and encodes a H. pylori GHP0 protein.
 CC The polypeptides can be used for preventing or treating Helicobacter
 CC infections, and gastroduodenal diseases associated with these
 CC infections, including acute, chronic, and atrophic gastritis, and peptic
 CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
 CC for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 SO Sequence 3166 BP; 1146 A; 448 C; 622 G; 950 T;

Query Match 77.9%; Score 14.8; DB 1; Length 3166;
 Best Local Similarity 88.9%; Pred. No. 74;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTGGCATTG 18
 Db 265 AGGCTATCTTGGCATTG 282

RESULT 14

ID X01362 standard; cDNA; 2085 BP.
 AC X01362;
 DT 16-APR-1999 (first entry)
 DE Nucleobase permease Ysp11 coding sequence.
 KW Ysp11; nucleobase permease; nucleobase binding affinity;
 KW drug screening assay; abnormal physiology; abnormal development;
 KW abnormal proliferation disease; cancerous condition;
 KW degenerative condition; ss.
 OS Mus sp.
 FH Key Location/Qualifiers
 FT CDS 48..1883
 FT /*tag= a
 FT /product= Ysp11
 FT misc-feature 366..1883
 FT /*tag= b
 FT /note= "Ysp11 form 2 created by alternative splicing"
 FT misc-feature 1515..1883
 FT /*tag= c
 FT /note= "Ysp11 form 3 created by alternative splicing"
 FT misc-feature 1659..1883
 FT /*tag= d
 FT /note= "Ysp11 form 4 created by alternative splicing"
 PN US5858707-A.
 PD 12-JAN-1999.
 PR 03-JUL-1996; 677049.
 PR 03-JUL-1995; US-000788.
 PR 03-JUL-1996; US-677049.
 PA (SCHE) SCHERING CORP.
 PI Bazan JF, Guimaraes MJ, McLanahan TK, Zlotnik A;
 DR WPI; 99-152103/13.
 DR P-PSDB; W73924.
 PT DNA encoding nucleobase permease polypeptides - useful for producing
 PT recombinant polypeptides
 PS Claim 10; Fig 2; 49pp; English.
 CC This sequence encodes the nucleobase permease of the invention.
 CC designated Ysp11. The nucleic acids are useful for producing recombinant
 CC polypeptides that have nucleobase binding affinity or nucleobase permease
 CC activity, for use in drug screening assays. The nucleobase permease can
 CC also be used in the treatment of conditions associated with abnormal
 CC physiology or development, including abnormal proliferation diseases such
 CC as cancerous conditions or degenerative conditions.
 SO Sequence 2085 BP; 364 A; 623 C; 548 G; 550 T;

Query Match 77.9%; Score 14.8; DB 1; Length 2085;
 Best Local Similarity 88.9%; Pred. No. 71;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GGCTTCTTGGCATTG 19
 Db 1699 GGCTTCTTGGCATTG 1716

RESULT 15

ID X20307 standard; DNA; 1907 BP.
 AC X20307;
 DT 04-MAY-1999 (first entry)
 DE Borrelia burgdorferi polynucleotide sequence #60.
 KW Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
 KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;


```

KW  Infection; diagnosis; characterisation; detection; ds.
US  Borrelia burgdorferi.
PN  W09858943-A1.
PD  30-DEC-1998.
PF  18-JUN-1998; 012764.
PR  03-SEP-1997; US-057483.
PR  20-JUN-1997; US-050359.
PR  22-JUL-1997; US-053344.
PR  22-JUL-1997; US-053377.
PA  (HUMA-) HUMAN GENOME SCI INC.
PA  (MED-) MEDIMATNE INC.
PI  Clayton R, Dougherty BA, Fraser C, Lathigra R, Smith HO,
PI  White OR:
DR  WPI; 99-081217/07.
PT  New isolated Borrelia burgdorferi nucleic acids - used to develop
PT  products for the detection, diagnosis, characterisation, prevention
PT  and therapy of infections, particularly Lyme disease
PS  Claim 1: Page 1042-1043; 1128pp; English.
CC  X20248 to X20402 represent polynucleotide sequences isolated from
CC  Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC  the detection, diagnosis, characterisation, prevention and therapy of
CC  Bb infections, e.g. Lyme disease. They can also be used for the
CC  production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC  to a family of motile, spiral-shaped bacteria called Spirochetes.
CC  Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC  endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC  Lyme disease.
SQ  Sequence 1907 BP; 615 A; 289 C; 164 G; 834 T;

Query Match 77.9%; Score 14.8; DB 1; Length 1907;
Best Local Similarity 88.9%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY  2 GCCTTCTTAGCCATCA 19
    ||||| |||||
DB  1690 GCCTTCTATAGCCATNA 1707

```

Search completed: September 13, 1999, 15:59:36
 Job time: 4803 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:56 ; Search time 1694.61 Seconds
(without alignments)
22.116 Million cell updates/sec

Title: US-09-325-095-25

Perfect score: 19
Sequence: 1 AGGCTTCTTATGCCATCA 19

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST.*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	89.5	444	26	W39661	W39661 zc19g05.r1
2	17	89.5	343	39	AA860212	AA860212 ak48d05.s
3	17	89.5	452	47	A1493815	A1493815 q294e09.x
4	17	89.5	422	47	A1493925	A1493925 q294f09.x
5	16	84.2	227	37	AA667104	AA667104 vq88f07.r
6	15.8	83.2	717	22	R81085	R81085 EST22e01 WA
7	15.8	83.2	320	23	H34794	H34794 EST112038 R
8	15.8	83.2	603	25	W12707	W12707 ma68d08.r1
9	15.8	83.2	576	31	AA277365	AA277365 va78h10.r
10	15.8	83.2	470	33	AA445601	AA445601 v158a12.r
11	15.8	83.2	464	34	AA462201	AA462201 vg73b01.r
12	15.8	83.2	514	34	AA466787	AA466787 vd90a12.r
13	15.8	83.2	502	36	AA641040	AA641040 nr72d10.s
14	15.8	83.2	331	42	A1154471	A1154471 ua02h12.r
15	15.8	83.2	521	43	A1101017	A1101017 EST210306
16	15.8	83.2	438	45	A1010636	A1010636 EST205087
17	15.8	83.2	446	45	A1327456	A1327456 ma68d08.y
18	15.8	83.2	452	45	A1373771	A1373771 q254f03.x
19	15.8	83.2	475	46	AA963348	AA963348 UI-R-E1-g
20	15.8	83.2	280	48	AT548121	AT548121 UI-R-C3-s
21	15.8	83.2	518	49	A1639435	A1639435 rx04153s
22	15.8	83.2	216	50	AA031440	AA031440 AV031440
23	15.8	81.1	371	31	AA290151	AA290151 VB35b02.r
24	15.4	81.1	277	39	AA834058	AA834058 cf25g04.s
25	15.4	81.1	266	47	A1462605	A1462605 vb35b02.x
26	15.4	81.1	378	47	AT508508	AT508508 vb35b02.y
27	15	78.9	347	26	W61987	W61987 md84c02.r1
28	15	78.9	472	29	AA161835	AA161835 ms17d03.r
29	15	78.9	211	29	AA169006	AA169006 ms35g04.r
30	15	78.9	439	30	AA212613	AA212613 mu44f12.r
31	15	78.9	493	34	AA472626	AA472626 vhl1b04.r
32	15	78.9	429	39	AA879952	AA879952 vm02d03.r
33	15	78.9	349	42	A1152256	A1152256 ud21c01.r
34	15	78.9	382	44	AU043395	AU043395 AU043395
35	15	78.9	288	46	A1414297	A1414297 md84c01.x
36	15	78.9	175	46	A1414750	A1414750 ma88e08.x
37	15	78.9	349	46	A1415238	A1415238 md14b07.x
38	15	78.9	519	48	AT607060	AT607060 ms17d03.y
39	14.8	77.9	299	20	T03898	T03898 seq2417.4HB
40	14.8	77.9	407	21	R09422	R09422 y122h09.r1
41	14.8	77.9	296	22	H03079	H03079 y146h12.r1
42	14.8	77.9	429	22	R31727	R31727 yb63e11.r1
43	14.8	77.9	231	22	R72272	R72272 y188e12.s1
44	14.8	77.9	476	24	H67434	H67434 yd52e01.r1
45	14.8	77.9	533	49	A1652703	A1652703 wb30g07.x

ALIGNMENTS

RESULT 1
LOCUS W39661
DEFINITION zc19g05.r1 Soares parathyroid tumor NBHPA Homo sapiens cDNA clone IMAGE:322808 5' similar to PIR:A35419 A35419 neutrophil protein -
ACCESSION W39661
NTD 9131370

VERSION	W39661.1	GI:1321370
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
AUTHORS	1 (bases 1 to 444) Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Holtzman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, B., Rohlfing, T., Soares, M., Tan, F., Trivaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.	
TITLE	The MashU-Merck EST Project	
JOURNAL	Unpublished (1995)	
COMMENT	On Nov 29, 1993 this sequence version replaced gi:430062.	
FEATURES	Source	
Source	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information. Insert Length: 705 Std Error: 0.00 Seq primer: mob.REGA+ET High quality sequence stop: 302. Location/Qualifiers 1..444 /organism="Homo sapiens" /db_xref="GDB:1260466" /db_xref="taxon:9606" /clone="IMAGE:322808" /clone_lib="Soares_parathyroid_tumor_MbHPA" /tissue_type="parathyroid tumor" /dev_stage="adult" /lab_host="DH10B (ampicillin resistant)" /note="Organ: parathyroid gland; Vector: pVT73D (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; 1st strand cDNA was primed with a Not I - 15'- TGTTCACATCTGAAGTGGAGCGCCGCCACCATTTTTTTTTTTTTTTTTTTT T-3'), double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pVT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Felima Ronaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."	
BASE COUNT	161 a 67 c 131 g 83 t 2 others	
ORIGIN		
Query Match	89.5%; Score 17; DB 26; Length 444;	
Best Local Similarity	100.0%; Pred. No. 50;	
Matches 17; Conservative	0; Mismatches	0; Indels
0y	1 AGGCTTCTTTAGCCAT 17	0; Gaps 0;
Db	392 AGGCTTCTTTAGCCAT 408	
RESULT 2		
LOCUS	AA860212 343 bp mRNA EST 04-JAN-1999	
DEFINITION	ak48d05.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1409193	
ACCESSION	AA860212	
VERSION	1.1 GI:2954207	

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 343)
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997) On Sep 12, 1996 this sequence version replaced gi:1400930.
CONTACT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
CDNA LIBRARY	CDNA Library Arrayed by: Greg Lennon, Ph.D.
SEQUENCING	DNA Sequencing by: Washington University Genome Sequencing Center
CLONE DISTRIBUTION	clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www.bio.llnl.gov/bdrp/image/image.html
FEATURES	Insert Length: 796 Std Error: 0.00 Seq Primer: -40ml3 fwd. ET from Amersham High quality sequence stop: 315. Location/Qualifiers 1. 343 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:1409193" /clone_lib="Soares_testis_NHT" /sex="male" /lab_host="DH10B" /note="vector: pT73D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc, and primed with a Not I - oligo(dT) primer [5', TGTTACCAATCGAAGTGGAGCGCGCCCAATTTTTTTTTT 3'] . Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cots, and was constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT	99 a 67 c 55 g 122 t
ORIGIN	
Query Match	89.5%; Score 17; DB 39; Length 343;
Best Local Similarity	100.0%; Pred. No. 51;
Matches 17; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AGCGTTCTTAGCCAT 17 	
Db 260 AGCGTTCTTAGCCAT 276	
RESULT 3	
LOCUS A1493915	
DEFINITION	A1493915 452 bp mRNA EST 18-MAR-1999
IMAGE:2042248 3', mRNA sequence.	CG946e09.x1 Soares_pregnant_uterus_NbH9D Homo sapiens cDNA clone
ACCESSION	A1493915
NID	94394918
VERSION	A1493915.1
KEYWORDS	GI:4394918
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
TITLE	Eutheria; Primates; Catarrhini; Homnidae; Homo.
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .
COMMENT	1 (bases 1 to 452)

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On May 18, 1998 this sequence version replaced gi:3137579.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 517 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 429.

FEATURES

SOURCE

1. 452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="14"
/clone_lib="Soares_pregnant_uterus_JbHPV"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer 15',
AACTGGAGAAATTCGCGCCCTTTTCTTTTCTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 125 a 105 c 79 g 143 t

ORIGIN

Query Match 89.5%; Score 17; DB 47; Length 452;
Best Local Similarity 100.0%; Pred. No. 50;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTTAGCCAT 17
|||||
Db 252 AGGCTTCTTTAGCCAT 268

RESULT 4
A1493925 422 bp mRNA EST 18-MAR-1999
LOCUS q204f09.x1 Soares_pregnant_uterus_NbHPV Homo sapiens cDNA clone
DEFINITION IMAGE:2042249 3', mRNA sequence.
ACCESSION A1493925
NID 94394928
VERSION A1493925.1 GI:4394928
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 422)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
Tumor Gene Index
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Unpublished (1997)
JOURNAL On May 18, 1998 this sequence version replaced gi:3137589.
COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 516 Std Error: 0.00
Seq primer: -400p from Gibco
High quality sequence stop: 416.

FEATURES

SOURCE

Location/Qualifiers
1. 422
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Soares_pregnant_uterus_NbHPV"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: uterus; Vector: pT73-Pac; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dt) primer 15',
AACTGGAGAAATTCGCGCCCTTTTCTTTTCTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."

BASE COUNT 118 a 94 c 72 g 138 t

ORIGIN

Query Match 89.5%; Score 17; DB 47; Length 422;
Best Local Similarity 100.0%; Pred. No. 50;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTTAGCCAT 17
|||||
Db 254 AGGCTTCTTTAGCCAT 270

RESULT 5

LOCUS

AA667104 227 bp mRNA EST 19-NOV-1997
v988f07.r1 Knowles Solter mouse blastocyst B3 Mus musculus cDNA
clone IMAGE:1109413 5' similar to gb:MI8194 Mouse fibronectin
(MOUSE);, mRNA sequence.

ACCESSION AA667104
NID 92625805
VERSION AA667104.1 GI:2625805
KEYWORDS EST.

SOURCE

house mouse.
Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 227)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, J.,
Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)
On May 9, 1995 this sequence version replaced gi:803003.

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:607581

Trace considered overall poor quality
High quality sequence stop: 1.

FEATURES

SOURCE

Location/Qualifiers
1. 227
/organism="Mus musculus"
/strain="B6D2 F1/7"
/db_xref="taxon:10090"
/clone_lib="IMAGE:1109413"

```

/clone_lib="Knowles Solter mouse blastocyst B3"
/issue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pSPORT; Site: 1: NotI;
Site: 2: SalI; cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI(dT):
5'-CGGTGACCGTCGACCGTTTATTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: B1 (larger
inserts) and B3."
BASE COUNT      59 a      54 c      38 g      76 t
ORIGIN

Query Match      84.2%; Score 16; DB 37; Length 227;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 GCCTTCTTAGCATC 18
        |||||||
Db      12 GCCTTCTTAGCATC 27

RESULT 6
R81085/c      R81085      717 bp      mRNA      EST      12-JUN-1996
LOCUS      EST22e01 WATM1 Homo sapiens cDNA clone 22e01, mRNA sequence.
DEFINITION      R81085
ACCESSION      R81085
NID      9857688
VERSION      R81085.1 GI:857688
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1 (bases 1 to 717)
AUTHORS      Boulland,F.
TITLE      Study of expressed sequences tags in adipose tissue 1995
JOURNAL      Unpublished (1995)
COMMENT      *on May 8, 1995 this sequence version replaced gi:801224.

Contact: Frederic Boulland
Centre de Recherche sur l'Endocrinologie moleculaire et le
Developpement
CNRS
9, Rue Jules Hetzel, Meudon Bellevue, 92190 France
Tel: 33 1 45 07 52 87
Fax: 33 1 45 07 58 90
Email: boulland@infobiogen.fr
automatic cycle sequencing of PCR amplified insert
Insert Length: 851 Std Error: 0.00
High quality sequence stop: 290.
Location/Qualifiers
1..717
/organism="Homo sapiens"
/strain="Caucasian"
/db_xref="taxon:9606"
/clone_lib="22e01"
/notes="Vector: lambda gt11; Site: 1: EcoRI; White adipose
tissue, sub cutaneous, adult, female. Purification of
polya mRNA, first strand priming with random
oligonucleotides. Ligation of adaptors EcoRI/NotI, cloning
in EcoRI site of lambda gt11."
BASE COUNT      196 a      172 c      192 g      157 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 22; Length 717;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 AGGCTTCTTAGCATCA 19
        |||||||
Db      252 AGGCTTCTTAGCATCA 234

RESULT 7
H34794/c      H34794      320 bp      mRNA      EST      13-MAR-1998
LOCUS      EST112038 Rat PC-12 cells, NGF-treated (9 days) Rattus sp. cDNA 5'
DEFINITION      end, mRNA sequence.
ACCESSION      H34794
NID      9980211
VERSION      H34794.1 GI:980211
KEYWORDS      EST.
SOURCE      Rattus sp.
ORGANISM      Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE      1 (bases 1 to 320)
AUTHORS      Lee,N.H., Weinstein,K.G., Kirkness,E.F., Earle-Hughes,J.A.,
Fulmer,R.A., Marmaras,S., Glodex,A., Cocayne,J.D., Adams,M.D.,
Kerlavage,A.R., Fraser,C.M. and Venter,J.C.
Comparative expressed-sequence-tag analysis of differential gene
expression profiles in PC-12 cells before and after nerve growth
factor treatment
Proc. Natl. Acad. Sci. U.S.A. 92, 8303-8307 (1995)
95396786

JOURNAL
MEDLINE
COMMENT

Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
For clone availability please contact the TIGR database
(tdbinfo@db.tigr.org)
Insert Length: 855 Std Error: 0.00
Seq primer: M13 Reverse
High quality sequence stop: 1.
Location/Qualifiers
1..320
/organism="Rattus sp."
/db_xref="ATCC (inhost):2066129"
/db_xref="taxon:10118"
/clone_lib="Rat PC-12 cells, NGF-treated (9 days)"
/notes="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:
XhoI; poly(A)+ RNA was purified from 9-day NGF treated
PC12 cells. cDNA was constructed using an oligo-dT primer
and directionally cloned using the Lambda Zap II Vector
Kit by Stratagene"
BASE COUNT      106 a      61 c      74 g      79 t
ORIGIN

Query Match      83.2%; Score 15.8; DB 23; Length 320;
Best Local Similarity 89.5%; Pred. No. 2.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY      1 AGGCTTCTTAGCATCA 19
        |||||||
Db      22 AGGCTTATTTAGCATGA 4

RESULT 8
W12707      W12707      603 bp      mRNA      EST      02-OCT-1997
LOCUS      W12707
DEFINITION      ma68d08.r1 Soares mouse p3NM19.5 Mus musculus cDNA clone
IMAGE:315855 5' similar to SW:PEPI_RAT P07734 BRAIN SPECIFIC
POLYPEPTIDE PEP-19 ;, mRNA sequence.
ACCESSION      W12707
NID      91286817
VERSION      W12707.1 GI:1286817

```

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 603)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 21, 1992 this sequence version replaced g1:279349.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI: 206471
Seq primer: mob.REGA+ET
High quality sequence stop: 459.
Location/Qualifiers
1. 603
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:315855"
/clone_id="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTACCAATCTGAAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT 156 a 165 c 145 g 137 t
ORIGIN
Query Match 83.2%; Score 15.8; DB 25; Length 603;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGCTTCTTAGCCATCA 19
DB 299 AGGCTCTCTTAGCCATCA 317
RESULT 9
LOCUS AA277365 576 bp mRNA EST 01-APR-1997
DEFINITION va78h10.r1 Soares mouse NML Mus musculus cDNA clone IMAGE:737539 5'
similar to SW:PE19_HUMAN P48539 BRAIN SPECIFIC POLYPEPTIDE PEP-19.
; mRNA sequence.
ACCESSION AA277365
NID 91917845
VERSION AA277365.1 GI:1917845
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 576)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Dec 30, 1996 this sequence version replaced g1:1529530.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:45387
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 497.
Location/Qualifiers
1. 576
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:737539"
/clone_id="Soares mouse NML"
/issue_type="Liver"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5,
TGTACCAATCTGAAAGTGGAGCGCGCATTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adapters
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
constructed and normalized by Bento Soares and M. Fatima
Bonaldo."
BASE COUNT 138 a 194 c 136 g 108 t
ORIGIN
Query Match 83.2%; Score 15.8; DB 31; Length 576;
Best Local Similarity 89.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 AGGCTTCTTAGCCATCA 19
DB 474 AGGCTCTCTTAGCCATCA 492
RESULT 10
LOCUS AA445601 470 bp mRNA EST 03-JUN-1997
DEFINITION v158a12.r1 Soares mouse NDMH Mus musculus cDNA clone IMAGE:847966
5' similar to SW:PE19_HUMAN P48539 BRAIN SPECIFIC POLYPEPTIDE
PEP-19. ; mRNA sequence.
ACCESSION AA445601
NID 92158284
VERSION AA445601.1 GI:2158284
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 470)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project

JOURNAL Unpublished (1996)
COMMENT On May 9, 1995 this sequence version replaced g1:802535.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:510018
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 466.
Location/Qualifiers

FEATURES

source

1. 470
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:847966"
/clone_1lb="Soares mouse NBMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCATCTGAGTGGAGCGCGCGGAAGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRTT3 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 128 a 127 c 115 g 100 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 33; Length 470;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
||||| ||||| ||||| |||||
Db 306 AGGCTCTCTTAGCCATCA 324

RESULT 11
LOCUS AA462201 464 bp mRNA EST 10-JUN-1997
DEFINITION v973b01.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:871561
ACCESSION AA462201
VERSION 92187092
KEYWORDS AA462201.1 GI:2187092
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 464)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced g1:1400786.
Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGI:511041
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 422.
Location/Qualifiers

FEATURES

source

1. 464
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:871561"
/clone_1lb="Soares mouse NBMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pRTT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCATCTGAGTGGAGCGCGCGGAAGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pRTT3 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldo."

BASE COUNT 110 a 126 c 112 g 116 t
ORIGIN

Query Match 83.2%; Score 15.8; DB 34; Length 464;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 AGGCTTCTTTAGCCATCA 19
||||| ||||| ||||| |||||
Db 73 AGGCTCTCTTAGCCATCA 91

RESULT 12
LOCUS AA466787 514 bp mRNA EST 11-JUN-1997
DEFINITION v990a12.r1 Soares mouse NBMH Mus musculus cDNA clone IMAGE:807838
ACCESSION AA466787
VERSION 92192927
KEYWORDS AA466787.1 GI:2192927
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 514)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 25, 1995 this sequence version replaced g1:637840.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:484182
Seq primer: -26ml3 rev2 ET from Amersham
High quality sequence stop: 510.

FEATURES

Source

Location/Qualifiers
1. 514
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:807838"
/clone_1lb="Soares mouse NDMH"
/sex="male"
/tissue_type="heart"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTCACATCTGAGAGGAGCGCGCGACAAAGTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Patima
Bonalido."

BASE COUNT

117 a 155 c 110 g 132 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 34; Length 514;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTAGCCATCA 19

Db 149 AGGCTCTTAGCCATCA 167

RESULT 13
AA641040/c 502 bp mRNA EST 27-OCT-1997
LOCUS nt72d10.s1 NCI-CGAP-Pr24 Homo sapiens cDNA clone IMAGE:1173523 3'
DEFINITION similar to gb:D13748 EUKARYOTIC INITIATION FACTOR 4A-1 (HUMAN);,
mRNA sequence.

ACCESSION
AA641040
NID 92566290

VERSION
AA641040.1 GI:2566290
KEYWORDS
EST.

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 502)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL
On Apr 14, 1993 this sequence version replaced gi:692623.

COMMENT

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,
Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 146.

FEATURES

Source

Location/Qualifiers
1. 502
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1173523"
/clone_1lb="NCI-CGAP-Pr24"
/tissue_type="invasive tumor (cell line)"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: prostate; Vector: Bluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Invasive prostate tumor cell line (HPV
immortalized). 5' adaptor sequence: 5' GAATTCGGCAGAG 3'
3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'
Average insert size: 1.0 Kb."

BASE COUNT

120 a 125 c 120 g 137 t

ORIGIN

Query Match 83.2%; Score 15.8; DB 36; Length 502;
Best Local Similarity 89.5%; Pred. No. 2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 AGGCTTCTTAGCCATCA 19

Db 49 AGGTTTCTTAGTCATCA 31

RESULT 14
A1154471 331 bp mRNA EST 30-SEP-1998
LOCUS us02h12.r1 Soares 2NBMT Mus musculus cDNA clone IMAGE:1345607 5',
DEFINITION mRNA sequence.
A1154471
ACCESSION
g3682940
NID A1154471.1 GI:3682940
VERSION
A1154471.1 GI:3682940
KEYWORDS
EST.

SOURCE

ORGANISM

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 331)
Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMNI Mouse EST Project
Unpublished (1996)
On Nov 29, 1993 this sequence version replaced gi:636069.

JOURNAL

On Nov 29, 1993 this sequence version replaced gi:636069.

COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMNI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:694399
Seq primer: -26ml3 rev2 ET from Amersham
High quality sequence stop: 207.

FEATURES

Source

Location/Qualifiers
1. 331
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="3"
/clone="IMAGE:1345607"
/clone_1lb="Soares 2NBMT"

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:58 ; Search time 539.84 Seconds

(without alignments)
117.824 Million cell updates/sec

Title: US-09-325-095-27

Perfect score: 20

Sequence: 1 GGNGAGATTCGNCARGA 20

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*
1: gb_da1.*
2: gb_da2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pi1.*
8: gb_pi2.*
9: gb_pi3.*
10: gb_pi4.*
11: gb_pi5.*
12: gb_pi6.*
13: gb_pi7.*
14: gb_pi8.*
15: gb_pi9.*
16: gb_pi10.*
17: gb_pi11.*
18: gb_pi12.*
19: gb_pi13.*
20: gb_pi14.*
21: gb_pi15.*
22: gb_pi16.*
23: gb_pi17.*
24: gb_pi18.*
25: gb_pi19.*
26: gb_pi20.*
27: gb_pi21.*
28: gb_pi22.*
29: gb_pi23.*
30: gb_pi24.*
31: gb_pi25.*
32: gb_pi26.*
33: gb_pi27.*
34: gb_pi28.*
35: gb_pi29.*
36: gb_pi30.*
37: gb_pi31.*
38: gb_pi32.*
39: gb_pi33.*
40: gb_pi34.*
41: gb_pi35.*
42: gb_pi36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	16	80.0	3207	3	BOVPHOS3KN	M93252 Bovine phosp

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	FEATURES	Source
1	BOVPHOS3KN	Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete	M93252	1	GI:163519	Phosphatidylinositol 3-kinase.	Bos taurus	1 (bases 1 to 3207)	Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R., Panayiotou, G., Ruiz-Iarrea, F., Thompson, A.S., Totty, N.F., Hsuan, J.J., Courtneidge, S.A., Parker, P.J., and Waterfield, M.D.	Phosphatidylinositol 3-kinase: Structure and expression of the 110 kd catalytic subunit	Cell 70, 419-429 (1992)	92354059	Location/Qualifiers	
2	16	80.0	3808	3	SSP120									
3	16	80.0	3452	4	AF001076									
4	16	80.0	3944	4	GGU95088									
5	16	80.0	3498	5	A37232									
6	16	80.0	4134	5	A50263									
7	16	80.0	4137	5	A50265									
8	16	80.0	339	8	CEU97661									
9	16	80.0	2200	8	CRU97662									
10	16	80.0	2634	8	SPU32583									
11	16	80.0	5397	10	HSIDEM									
12	16	80.0	3424	10	HSP13K									
13	16	80.0	2990	10	HSP17R1									
14	16	80.0	3207	10	HSU79143									
15	16	80.0	3213	10	S67334									
16	16	80.0	3207	12	MMU03279									
17	16	80.0	3213	12	RNO012482									
18	16	80.0	2752	12	RNO6710									
19	16	80.0	3389	17	AF001075									
20	16	80.0	135259	34	AC005018									
21	16	80.0	17370	36	CELB0025									
22	16	80.0	2837	36	CEVPS34									
23	16	80.0	5577	36	DDU23477									
24	16	80.0	3165	36	DMP13K59F									
25	16	80.0	188288	37	AC005639									
26	15	75.0	115602	9	HS1118D24									
27	14.8	74.0	2612	1	CGDNAADP									
28	14.8	74.0	24800	2	SC9B1									
29	14.8	74.0	11925	2	AE001013									
30	14.8	74.0	3331	3	BOVNT0XSY									
31	14.8	74.0	3859	3	BTU88531									
32	14.8	74.0	1997	4	AF001413									
33	14.8	74.0	1612	5	AS8887									
34	14.8	74.0	2987	7	SOYPI3K5X									
35	14.8	74.0	2760	7	U10669									
36	14.8	74.0	120787	8	AC002292									
37	14.8	74.0	3471	9	AB005910									
38	14.8	74.0	3261	9	HSR011121									
39	14.8	74.0	3599	9	HSR011122									
40	14.8	74.0	3599	9	HSR011123									
41	14.8	74.0	4880	9	HSC2P13K1									
42	14.8	74.0	7654	10	HSC2P13K									
43	14.8	74.0	3868	10	HSP110DEL									
44	14.8	74.0	5296	10	HSP13K1NA									
45	14.8	74.0	62980	35	AC007771									

ALIGNMENTS

RESULT 1
BOVPHOS3KN
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
FEATURES
Source

BOVPHOS3KN 3207 bp mRNA
Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete
M93252
1 (bases 1 to 3207)
GI:163519
Phosphatidylinositol 3-kinase.
Bos taurus cDNA to mRNA.
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R.,
Panayiotou, G., Ruiz-Iarrea, F., Thompson, A.S., Totty, N.F.,
Hsuan, J.J., Courtneidge, S.A., Parker, P.J., and Waterfield, M.D.
Phosphatidylinositol 3-kinase: Structure and expression of the 110
kd catalytic subunit
Cell 70, 419-429 (1992)
92354059
Location/Qualifiers
1. 3207

SHAGLSNRLANDNEIRENDKQALRAICTRDPISSETEDEKDFLMSHRYCWTIPILP
KLLSVKNSRDEVAQMYCLVADWPIKPEQAMELLDCNDPPMVRGFAVRLERYLT
DDKLSQVLIQVQVAKYEQYLDNLLVRELKALTNORIGHEFWHLKSEMKNVSO
REGLLSEYCRAGMYLKLHNOVAMEKLIJTLIOEKDETOYKOMEFLVOMR
RPDEFDALOGFLSPNPAHOLGNLWLECRIMSSAKRPLUMWENPDIMSEHPQNE
IIFKNGDRLRDMLELOITRIRMENTWOGDRLRMLPRGCLSIGCVGLIEVRNHT
IMQJCKGGLGALDFNSHTLHOMLKNKGITDIAIDLFTRSCAGTCVATFIIIGIG
DRHNSINWAKDQGLFIDFHFLOHKKRKEGYKREPVLTDFLIVISGADECT
KTREPEROEMCYKAYLIAIROMANLFINFMSMLGSGMPELOSFDDIAYIRKTLALDK
TEOEALFEYFMKOMDANHGGWTKMDWIFHTIKOHALN"

BASE COUNT 1147 a 626 c 722 g 1003 t
ORIGIN
Query Match 80.0%; Score 16; DB 5; Length 3498;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GNGAYGAYTTCGNCARGA 20
DB 2410 GGGATGATTCGCGCAGA 2429

RESULT 6
LOCUS A50263 4134 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 1 from Patent WO9612024.
ACCESSION A50263
NID 92303320
VERSION A50263.1 GI:2303320
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4134)
Stoyanov, B., Hancock, T. and Wetzker, R.
TITLE CLONING, EXPRESSION AND CHARACTERISATION OF A NOVEL FORM OF
PHOSPHATIDYLINOSITOL-3-KINASE
JOURNAL Patent: WO 9612024-A 1 25-APR-1996;
MAX PLANCK GESELLSCHAFT (DE)
COMMENT Other publication DE 4445562 960404.
FEATURES
source 1. 4134
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 1127 a 999 c 998 g 1010 t
ORIGIN
Query Match 80.0%; Score 16; DB 5; Length 4134;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GNGAYGAYTTCGNCARGA 20
DB 2820 GGTGATGATTCGCGCAGA 2839

RESULT 7
LOCUS A50265 4137 bp DNA PAT 07-MAR-1997
DEFINITION Sequence 3 from Patent WO9612024.
ACCESSION A50265
NID 92303321
VERSION A50265.1 GI:2303321
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 4137)
Stoyanov, B., Hancock, T. and Wetzker, R.

TITLE CLONING, EXPRESSION AND CHARACTERISATION OF A NOVEL FORM OF
PHOSPHATIDYLINOSITOL-3-KINASE
JOURNAL Patent: WO 9612024-A 3 25-APR-1996;
MAX PLANCK GESELLSCHAFT (DE)
COMMENT Other publication DE 4445562 960404.
FEATURES
source 1. 4137
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 1127 a 999 c 1001 g 1010 t
ORIGIN
Query Match 80.0%; Score 16; DB 5; Length 4137;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GNGAYGAYTTCGNCARGA 20
DB 2823 GGTGATGATTCGCGCAGA 2842

RESULT 8
LOCUS CEN97661 339 bp mRNA PLN 26-MAY-1998
DEFINITION Chlamydomonas eugametos phosphatidylinositol 3-kinase mRNA, partial
cds.
ACCESSION U97661
NID 92109286
VERSION U97661.1 GI:2109286
KEYWORDS
SOURCE Chlamydomonas eugametos.
ORGANISM Chlamydomonas eugametos
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadales; Chlamydomonadales.
REFERENCE 1 (bases 1 to 339)
Molendijk, A.J. and Irvine, R.F.
TITLE Inositolide signalling in Chlamydomonas: characterization of a
phosphatidylinositol 3-kinase gene
JOURNAL Plant Mol. Biol. 37 (1), 53-66 (1998)
MEDLINE 98281574
REFERENCE 2 (bases 1 to 339)
Molendijk, A.J. and Irvine, R.F.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1997) Pharmacology, University of Cambridge,
Tennis Court Road, Cambridge, United Kingdom
FEATURES
source 1. 339
/organism="Chlamydomonas eugametos"
/db_xref="taxon:3053"
/note="YPS34 homolog"

CDS
CDS
/product="phosphatidylinositol 3-kinase"
/protein_id="AAC50016.1"
/db_xref="PID:92109287"
/db_xref="GI:2109287"
/translation="GDDLRQDQVLTQLFALMDRLRLRDVNLDRITPPYHLATSPDGL
VOFIPSVAAVAQVGYGDLISYLRHFYDPDENATFHVQAOVLDTFVSCAGYCVITY
LLGVGDRHLN"

BASE COUNT 68 a 102 c 94 g 75 t
ORIGIN
Query Match 80.0%; Score 16; DB 8; Length 339;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
QY 1 GNGAYGAYTTCGNCARGA 20
DB 1 GGGATGATTCGCGCAGA 20

RESULT 9
CRU97662 2200 bp mRNA PLN 26-MAY-1998
LOCUS Chlamydomonas reinhardtii phosphatidylinositol 3-kinase mRNA,
partial cds.
DEFINITION
ACCESSION U97662
NID 92109288
VERSION U97662.1 GI:2109288
KEYWORDS
SOURCE Chlamydomonas reinhardtii.
ORGANISM Chlamydomonas reinhardtii.
REFERENCE 1 (bases 1 to 2200)
AUTHORS Molendijk,A.J. and Irvine,R.F.
TITLE Inositolide signalling in Chlamydomonas: characterization of a phosphatidylinositol 3-kinase gene
JOURNAL Plant Mol. Biol. 37 (1), 53-66 (1998)
MEDLINE 98281574
REFERENCE 2 (bases 1 to 2200)
AUTHORS Molendijk,A.J. and Irvine,R.F.
TITLE Direct Submission
JOURNAL Submitted (15-APR-1997) Pharmacology, University of Cambridge, Tennis Court Road, Cambridge, United Kingdom
FEATURES
source location/Qualifiers
1..2200
/organism="Chlamydomonas reinhardtii"
/strain="CW-15"
/db_xref="taxon:3055"
<1..2199
/note="Vps34 homolog"
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="AAC50017.1"
/db_xref="PID:92109289"
/db_xref="GI:2109289"
/translation="ESGAGAGSGPRGGAITLVNPEVGRDPAPAEIKAKKLRGICQRCV
LDRDKPSEERRRQATILKLPNKPPLHAERBALWFRASLTQDTRLTFFLOCVDM
SDATEAKOAAELMRQWAPISADALELSPFRESVNRQAVQALQAAVEVELAYTL
QVQALREADDTGSRGAPLVORAAEDPALAVQHWLFCGLDPGGLRABALHTL
LDDTLAGGPRSPGILLADTIRQVDVYKQKAVIRELSSKIKAAKATEBLREMSV
TGAGGELALVPLIPDLPHYIAGIMPRECSVFSKAMPLRISFRTAAAPMWRAPPP
PQPHNTHNPHPHFPHGLHGHAPAGASHLSGSGSSSSGAYHSPGPGCGCGGG
WPPDPPLALPTRTGILPAPGAPDPAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG
ASRGAGGAGAPSGPAPVLEGRCCIIRKGGDLQODHFVQLQIMDMRMKREN
LDLRMTPKVLPSTSDGLVEFVSVPASAVLAERKITHREFLATQADAGGPFGRAR
AVETFRSCAGYCWTVILIGDRLDMLTHGGRFLHIDFGYILGADTKPFPMPK
LCKEMTEMGSGSEYKOPRMVACAEANIRKADLTLSFHLMAGASTAIRADPE
KAILKLDKFLDDEDAVEMOGLINESATALLPDLVETTHRAQYWR"
BASE COUNT 347 a 775 c 772 g 306 t
ORIGIN
Query Match 80.0%; Score 16; DB 8; Length 2200;
Best Local Similarity 65.0%; Pred. NO. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

RESULT 11
HSD1EM 5397 bp mRNA PRI 07-OCT-1996
LOCUS H.sapiens mRNA for phosphatidylinositol 3 kinase gamma.
DEFINITION
ACCESSION X83368
NID 91507821
VERSION X83368.1 GI:1507821
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 5397)
AUTHORS Stoyanov,B., Voljina,S., Hanck,T., Rubio,I., Loubtchenkov,M., Malek,D., Stoyanova,S., Vanhaesebroeck,B., Dhand,R., Nuernberg,B., Gierschik,P., Seedorf,K., Hsuan,J.J., Waterfield,M.D. and
Eukaryota: Fungi; Ascomycota; Archiascomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;
Schizosaccharomyces.
REFERENCE 1 (bases 1 to 2634)
AUTHORS Takegawa,K., Denald,D.B. and Emr,S.D.
TITLE Schizosaccharomyces pombe Vps34p, a phosphatidylinositol-specific PI 3-kinase essential for normal cell growth and vacuole morphology
JOURNAL J. Cell. Sci. 108 (Pt 12), 3745-3756 (1995)
MEDLINE 96360252
REFERENCE 2 (bases 1 to 2634)
AUTHORS Takegawa,K., Denald,D.B. and Emr,S.E.
TITLE Direct Submission
JOURNAL Submitted (28-JUL-1995) Scott Emr, Div. of Cellular and Molecular Medicine, University of California at San Diego School of Medicine, 9500 Gilman Drive, San Diego, CA 92093-0668, USA
FEATURES
source location/Qualifiers
1..2634
/organism="Schizosaccharomyces pombe"
/db_xref="taxon:4896"
57..2462
/gene="Spvps34"
57..2462
/gene="Spvps34"
/note="Spvps34"
/codon_start=1
/product="Spvps34p"
/protein_id="AAC49133.1"
/db_xref="PID:929997"
/db_xref="GI:929997"
/translation="MDRLVPSYCPSSKVTARFLVYKFCFIEYODSOECICTIOFSN
ESGSLMOKCTYPSKIPKSLPTELKSTHEMLDQGVTSLSNAKFVSAKPSRN
DEEVEYEGGCTYRLFDENNLKRGDLKIPQTSKEIKTSPSLEQVKEINRLKG
LLIQLOQGVPSVMWLDISFGKIKDRSRHMSLVTPILYLFQFSFVYQRSY
PKSENRVYSSFDLELNDSPAEKHLRYSONRGJLDDKPKSKIRLELSISY
PSESLLEEDDLTKRFFYLRKMKATFKELKVWTDSEVNOALSLDSMEIDI
DDALELSPSPVHPKVAAYAVRLETDSNEELLYLQVQALRYNDPISDERFOPS
PLAFVYNRAISSSTINDUXWYLYVEIEDEPVSKLTSVWFLQKELSVBERLIR
ETLSAQAKFEKLRISKSVQSEFGTLKIKIETKVLLEDKHLDLDFHALPLDLDS
VNIIGILPDCTVYKSTWOPRLRFLKQDSSKYPPIKRNDDLDQDLVQIITLDMK
LKKEKEDLHKPRLIATPGTPHGAQFVSQKTIATIAEHGSLVLAIRENNDDGL
NSAYGIDPAMDVYRSQAGYCYITVLLGVGDHLNLLITKGFHFDGFIILGR
DKPIFSAMLSKMEVSGMGNGYNRPFOQKRSYCYTFTALRSNNILNFSLMVA
NIPDIKDKKRYVYKVERCLQMSSEDAIKYFQILINDYSALFPQIIDRMHNLAY
MRS"
BASE COUNT 817 a 477 c 463 g 877 t
ORIGIN
Query Match 80.0%; Score 16; DB 8; Length 2634;
Best Local Similarity 65.0%; Pred. NO. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Wetzer, R.
Cloning and characterization of a G protein-activated human
phosphoinositide-3 kinase
JOURNAL
MEDLINE
95350661
REFERENCE
AUTHORS
JOURNAL
TITLE
2 (bases 1 to 5397)
Waterfield, M.D.
Direct Submission
Submitted (08-DEC-1994) M.D. Waterfield, Ludwig-Inst. for Cancer
Research, Courtauld Building, 91 Riding House Street, London, W1P
8BT, UK
REMARK
3 (bases 1 to 5397)
Waterfield, M.D.
Direct Submission
Submitted (23-AUG-1996) M.D. Waterfield, Ludwig-Inst. for Cancer
Research, Courtauld Building, 91 Riding House Street, London, W1P
8BT, UK
COMMENT
On Aug 26, 1996 this sequence version replaced gi:940513.
X83368 is homologous to M93252.
FEATURES
source
1. 5397
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="U937"
324..3629
/note="phosphatidylinositol 3 kinase gamma, p110 gamma,
activated by G protein alpha and betagamma subunits"
/codon_start=1
/product="idem"
/protein_id="CAA58284.1"
/db_xref="PID:664520"
/db_xref="PID:91507822"
/db_xref="GI:1507842"
/db_xref="SWISS-PROT:P48736"
/translation="MELENYKOPVILREDCNRRRRMRKPRSAASLSMELPIEFVLP
TSORCKSPETALHAGHNGVOMKAOYWLALSTVADYHRLGPHHLLIOLK
GOWEITDYQYVQTLDCRYKAKTHRSQQLHYVORHPSESOAFQQLALLIGD
VYDVSNHDELEFTRRGVLTFRMAEVSASRDKLAMPWVTSKPLPELTKIANNC
IFIVIHRTSQTIVSPDTPGALLOSFTMARKSLMDIPESQSEDFVLRVCGR
DELYGEPRIKFNQWRHCLKNGEIHVVLDPPALDEVKEMPLVDCTGYTG
HEQLTGHGDESEFTVSLMCDGRKFRVIRGIDIPVLRNDLVEYVANTOQCOV
LCORRSPKPEFEVLMWYLFESIKIDLPKALLINQIYCGKAPASLSKASAPS
SSKGRVLLYYVNLIDHRLIRGEVYLMQMOTSGGEOGSEFNADKLISANPD
KENSMSITLNDNCHPLAPKHPQIPDPEGDVRAEMPNQKOLEATIDPLNPL
TADEKELMHFYESLKHPRKAYPKLFSSVWQOGIYAKTYQLARREWDQALDVG
LTMQLDCNHFSDENRAIVOKLESLEDVLAHYLQVAVKFEFPHDSALARLK
RLRNKRIGHLEFLEFARSIAQSRHYQGFVILAYLRGCTAMLDHETQOVYEM
LOKVLDIRSLSAEYDVSSOYISOLKOKLEMLNSOLPESFRVYDPQLKGAALAE
KCKWASKKKPLLEKCAPALSNERTGIIIFKHGDILROMLIIQILRIMESMET
ESLDLCLEPYGISTGDKIGMIEIVKDATTKIOOSTYNGTGAIRDEVYLNHLEKS
PREEKFOAAVERFVSCAGYCVATFVLGIDPHNHNIMTETGNLFHIDFGHLLNFK
SFLGINKERVPVLTLPDLFVMTSGKSPHOKFODICVAVYALRHHHTLLILF
SMALTMGPOLTSKEDIERYIRDALTYGKNEEDAKRYFLDQIEVCBDKWTYQFNNFLH
LVLGKQGEKHSAA"

BASE COUNT 1534 a 1224 c 1207 g 1432 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 5397;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTTRCGNCARGA 20
IIIIIIIIIIIIIIIIIIII
Db 2823 GGATGATGATCTGGCCACAGA 2842

RESULT 12
LOCUS HSPH13K 3424 bp mRNA
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION 229090

NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
MEDLINE
95229146
REFERENCE
AUTHORS
JOURNAL
TITLE
2 (bases 1 to 3424)
Volinia, S.
Direct Submission
Submitted (16-DEC-1993) Stefano Volinia, Receptor Studies, Ludwig
Institute for Cancer, Research, 91 Riding House Street, London, W1P
8BT, UK
FEATURES
source
1. 3424
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3q26.3"
/cell_line="K562"
13..3219
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CAA82333.1"
/db_xref="PID:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-PROT:P42336"
/translation="MPRPSSGELMGILMPRIYVLCPLPNGMIVTLECLREATLVY
IKHELFKARKYPLHQLDSESYIFVSTVTEADREDEFTRRLCLRLPFLKVI
EPVGRREKILNREIFAIGPVCEDKPEVDKPEVDPRNTLNTCKEAVDRLNSPH
SRAMVYVPHYESPELPHKHYNLDRGQILVIVMIVSPNNDOKYTKIKNHCVPE
QVIAEATKTRSHLSSEOLKCLVLEQGYIILKVGCDYFLEKTPLSGYKTRSC
IMLGMPMLKMKAKESLSQLPMCCFTMPSTSRRIATPYMNETSTRSLWYNRL
RIKILCAVYVNLNRIDKIVRGYIHGGSPCLDNNVTOVPCSNPKNMMLNYDJY
IPDLPRARLLSLCSYSGRGAKEHCPLAMGWINLFDYDITVSGMALNLPVPH
GLEDLNLIGVTSNPKETPCLEEFPMFSSVVKFPMDSVIEEHANMSVREGFSY
SHAGLSVNDLRENDKRLAISTROPLSITTOEDKFLSHHRYCVTPEIILP
KLILSVKNSDENVDAQMLCYKMPPIPEBAMLLDGNIDPWRGFAVCKEYLIT
DDKSQITLQVLYAKTEQYLDNLDVRLKALFNORIGHFVHLKSEHANKTVSO
RFEGLESEACACGNYLKLNRQVEMAEKLIINDIDIKQERKDETQKQMKFLQNR
RPDMDALOGILSLNPAHQGNRLKECRIMSSAKRPLMNMENPDIMSLELQNNR
IIFKNGDRLDMGLTQIIRLMENIMONOGDLMLPYGCLSIDGAGLIEVNSHT
IMIOIQKGLKGLALOENSHLHOMLKKNKEIYDAIDLPFTRSCAGYCVATFVLGIG
DRHNSNIVKQDGLFHDPEGHFLDHKKKKRGRYRREVPYVLPDPLVYSKQAOECT
KTRFERFOEKCIVAYLAIROHANLFINLFSMGLSGPELOSFDDIAYIKRTLLDK
TEQVLEFEMQNDADHNGWTTKMDWTFHTIKQHALN"

BASE COUNT 1134 a 618 c 709 g 963 t
ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 3424;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTTRCGNCARGA 20
IIIIIIIIIIIIIIIIIIII
Db 2422 GGAGATGATTACGGCCACAGA 2441

RESULT 13
LOCUS HSPH13K 2990 bp mRNA
DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
ACCESSION 246973
NID 9987947
VERSION 246973.1 GI:987947
KEYWORDS phosphatidylinositol 3-kinase.

SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 887)
TITLE Volinia S., Dhand R., Vanhaesebroeck B., MacDougall L.R., Stein R.,
Zvelebil M.J., Dhand R., Panarecu C. and Waterfield M.D.
A human phosphatidylinositol 3-kinase complex related to the yeast
Vps34p-Vps15p protein sorting system
JOURNAL EMBO J. 14 (14), 3339-3348 (1995)
MEDLINE 95354652
REFERENCE 2 (bases 1 to 2990)
AUTHORS Volinia S.
TITLE Direct Submission
JOURNAL Submitted (21-DEC-1994) Stefano Volinia PhD, Receptor Studies,
Ludwig Institute For Cancer, Research, 91 Riding House Street,
London, W1P 8BT, UK
FEATURES
source Location/Qualifiers
1..2990
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="PIR-1"
/cell_line="TF-1, K51a"
/clone_1b="Lambda Zap TFI, K51a"
1..47
48..2711
/codon_start=1
/product="phosphatidylinositol 3-kinase"
/protein_id="CA87094.1"
/db_xref="PID:987948"
/db_xref="GI:987948"
/translation="MGEAEKFAHYISCDIDINVQLTIGSLGKREOKSYNAVLEDPML
KESGLIETCSLDLYTCQVFAEGKPSALPYRTSYKASTKMNENMLKLPKYPDLPR
NAOVALTIMDYVPGKAVPGTIVSLFGYGSROGMHLKVPNVEADGSEPTLP
GRSSLSLEDOMSLAKITFAHROGHVAVYDMDLPREIEIMNESVGRSSQYQYEF
GRRKVCDDKEXGIYKEDGESSPILSFELVYPPDQSELENVSKHNRLRS
LRSGPDHDKLPSPRDOKLIVSTPPSKPYTEREDOLWERTILINODKALIKIL
TSYIMDLPQAKALALILGNMPDVEDSLSSHYTNVTRVAVARLQADDDLD
LMTLSQVLAKEYENFDKNGDEPTKDSQSSGVNSGINSASVIAIDSSQITISPL
PSVSPSPASKTEVPDGENLEODICTFLISRAKNSSTLANLYMYVIVECEDPTQO
RDKTHTMYLVNWARFESQALLKGDKSVRVRSLAAOQTFVDRVLHLMKRAVQESGR
KKKNERLQALIGNEKMNLSVLEPLPLEPKYKIGIIPETATLTKSALMPQLPEK
TEBGKYPVTFKGGDDLRODQLIQIISLMDLKKRENDDIKLPTIKVLATSKHGM
QFISYVAEVLDTBGSIQEFKVIAPSENGPGISAEDVYKSCAGCVITTYLIG
VGDRLDNLVLTITGKLFHIDFGYIIGRPKPLPPMKLNKEVEMGQTSQYQEF
RKQCYATFLHRLRYSNLILNLFSLMVDNPIDIALPEPKTKVKKYQDKFRLLDSDEEAV
HYMOSLIDESVHALFAAVYEQIHKFAQYMKR"
BASE COUNT 2712 ..2990 566 c 678 g 797 t
ORIGIN
3'UTR
BASE COUNT 949 a 566 c 678 g 797 t
ORIGIN
Query Match 80.0%; Score 16; DB 10; Length 2990;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS 1 (bases 1 to 3207)
TITLE Stridivant S.M., Ahern J., Conroy R.R., Barnett S.F., Ledger L.M.,
O'Leary A. and Heimbrock D.C.
Catalytic Activity of the p110- α Subunit of Human
Phosphoinositide 3'-Hydroxylase is Required for Signal
Transduction
JOURNAL Bioorg. Med. Chem. (1996) In press
REFERENCE 2 (bases 1 to 3207)
AUTHORS Stridivant S.M., Ahern J., Conroy R.R., Barnett S.F., Ledger L.M.,
O'Leary A. and Heimbrock D.C.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1996) Cancer Research, Merck Research Labs,
Summerville Pike, West Point, PA 19486, USA
FEATURES
source Location/Qualifiers
1..3207
/organism="Homo sapiens"
/db_xref="taxon:9606"
/function="phospholipid kinase"
/product="phosphoinositide 3'-hydroxylase p110- α
subunit"
/protein_id="AAB39753.1"
/db_xref="PID:g1763626"
/db_xref="GI:1763626"
/translation="MPPRSGELMGHILMPRIIVECLLPNGMIVTLECLREATLIT
IKHELPKARKYPLHQLQDESSYIVFVSTOAEEREFDERLRCDDLPLPPLKVI
EPVGNPEKILIREIGFAGMPVCEPDMDQDEVDRENTILNVCENADLNSPH
SRMYVYPPNVSSPELPKHTINKDKQCIIVYIVYVSPNNDKQYTLKTHDCYPE
QVYAEIRKRTSMILSSQKLQVLEYGKYLIVCGDETFLEKYSQYKYLRSQ
IMGKPNMLMAKESLYSQLPMDCTFMSYSRRISTAPYNGKERTSLVINSAL
RIKILCATVANNIRIDIKIVRTGTYHSGELCNVNTQVPCSNPWNENLVNDY
IDPLPAAATLSTICSVGRKAKHEPCLAGNINLFEYTLVSGKALMLNVPVH
GLEDLNPICGVSNPNKPTCLTEPDMFSSVVRPMVSGCLSGDVGILVVRNHT
SHAGLSNRARNELENDKEDLKAISTDPSLSEITQKODLNSRHRYCVIPELP
KLILSVKMSRDEVAQMTCLVDMPPIREQAMELDCYTPDPVARGFVRLCEITLT
DKRLSQYLLQVQVLYKEQYLDNLVYRFLKRALNQRIGHEFFVHMLKEMNKTVSQ
RGILLESYCRACGMYLKLHNRQVEMETLNLITDLQEKDEQKVMKFLVEQR
RDEMDLQGFSPILPAHOLNLRLEECRINSKRPPLMNNENDISELFFORNE
IIFKNGDGLRQMLTIOIRIMENIMONGDILRMVPGCLSGDVGILVVRNHT
IMQIQCKGLKALQDNSTTHQWLKDKRKEITDAIDLFTRSCGCVVAFPIIGIG
DRHNSINWQKQFHLIDFGHLDKRRKFEIKRRVFFVILQDPLVITISGAQDCT
KREFEPEKQYKAYIALROHANEFINLSMLGSGMELOSFDIAYIRTLADLK
TDEALAEYMKQNDANHHGSGWTTKMDWIFHTIKOHALN"
BASE COUNT 1043 a 586 c 670 g 908 t
ORIGIN
Query Match 80.0%; Score 16; DB 10; Length 3207;
Best Local Similarity 65.0%; Pred. No. 1e+02;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

AUTHORS Hu P., Mondino A., Skolnik E.Y. and Schlessinger J.
TITLE Cloning of a novel, ubiquitously expressed human
phosphatidylinositol 3-kinase and identification of its binding
site on p85
JOURNAL Mol. Cell. Biol. 13 (12), 7677-7688 (1993)
MEDLINE 94067128
REMARK GenBank staff at the National Library of Medicine created this
entry [NCBI g140879] from the original journal article.
This sequence comes from Fig. 1.

FEATURES

source

1..3213

/organism="Homo sapiens"

/db_xref="taxon:9606"

gene

1..3213

/gene="phosphatidylinositol 3-kinase p110 beta isoform, PI
3-kinase p110 beta"

CDS

1..3213

/gene="phosphatidylinositol 3-kinase p110 beta isoform, PI
3-kinase p110 beta"/note="110 kda catalytic subunit; This sequence comes from
fig. 1; PI 3-kinase p110 beta"

/product="phosphatidylinositol 3-kinase p110 beta isoform"

/protein_id="AAB29081.1"

/db_xref="pid:9455760"

/db_xref="gi:455760"

/translation="MCFSEIMPAMADLIDIVANDSQTASDGSIPVDFLLPGIYITOL
EVPREATISYIKOMLMKOVHNPMPNLMIDISYMFACVNOAVYELEDETRRLCDV
RPLFLVLTWRSQDPKELSKIGVLIGKGLHEPDSLKDEPNEFRKMRKFSSEKI
LSVGLSWMDMKOTYPPHEPSPISLEDEKLYGSKLIYAVFENCQDVSFOVSPNM
NPIDYNEIAIOKRLTIHGKEDVSPDYVLOVSGRVEYFGDHPILIOFYIRNCVMNR
ALPHILVECCIKTKMEQEMATEAIAINRNSNPLPLPKRTIISHVENNPNPO
IYVGNKLNTEETKYVVRAGLPHGTTELCTIYSSEYSGKNDHINPELEFDINIC
DLPRARLCFAYAVLDKVKTKSTYQIRAKGVHVPVAVNTWVEDEK
QLRTGDIILHSWSPDELEKLNPMGTVOINPYENATALHVKFENKQPYYP
DKIIEKAAEIASDSANYSRGKFLPYLKEIILDRDPLSQCEEMDLITLRDGR
EIEFOSLPRILISIKWKLDEYVAQALLOIMPRLPREALEILDENYPDQYREYAV
GCLRMSDEELSOYLLOLVOLKYEPFLDCALSRILERAIGNRIRIGPLFMHLSEY
HIPAVSOGVILEAYICRGSVGHMKVLSKQYBALNKLKTLNSLILNNAVKLNRAKGE
AMHTLKQSAVREALSDLSPLNCPVILSELYVENCKYMSMKRPLMLVYNNKVFGE
SVGLFKNQDDELKQDMTLQMLRLMLKENGDLRLMYPGCLATGDRSGLIEVYST
SETIADIOLNSNVAAAAFNKDALNMLKEYNSGDDDLRAIEFTLSCAGCVASY
LGIGRHSNDINVKTKGOLFHDHFGHILGNFESKFGIKRERYPILTYDPIHVIOGK
TGNTEKPGFRQCCEDAYLILRRHGNLFITLPAIMLTAGLPELTSVKNDIQLKDSIAL
GKSEERALKQKQKDEALRESWTYVMAHIVAKDYRS"

BASE COUNT

979 a 612 c 704 g 918 t

ORIGIN

Query Match 80.0%; Score 16; DB 10; Length 3213;

Best Local Similarity 65.0%; Pred. No. 1e+02; Mismatches 13; Conservative 5; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20

DB 2419 GGTGATGATTACGACAGA 2438

Search completed: September 13, 1999, 15:55:59
Job time: 4587 sec

FT /note="PI3- kinase p110"

PN MO9321328-A.

LV 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDM-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R43341.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Claim 7; Fig 9: 146pp: English

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51156, Q59012-23 and Q57522-3.

SQ Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 80.0%; Score 16; DB 1; Length 3207;

Best Local Similarity 65.0%; Pred. No. 7.6;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGAYATTCGNCARGA 20

DB 2410 GGGGATGACTTACGACAGA 2429

RESULT 3

ID Q57522 standard; cDNA; 381 BP.

AC Q57522.

DT 12-APR-1994 (first entry)

DE Human PI3K-cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

PN MO9321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDM-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R46532.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Disclosure: Fig 20: 146pp: English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51156, Q59012-23 and Q57522-3.

SQ Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

CC glucose levels can be controlled using the kinase.

CC See also Q51156-6, Q59012-23 and Q57523.

SQ Sequence 381 BP; 116 A; 78 C; 85 G; 102 T;

Query Match 80.0%; Score 16; DB 1; Length 381;

Best Local Similarity 65.0%; Pred. No. 5.7;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGAYATTCGNCARGA 20

DB 1 GGGAGAGACTTCGACAGCA 20

RESULT 4

ID Q57523 standard; cDNA; 393 BP.

AC Q57523.

DT 12-APR-1994 (first entry)

DE Human PI3K-cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

PN MO9321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDM-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R46533.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PS activity, useful for controlling cell proliferation

PS Disclosure: Fig 21: 146pp: English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also Q51156-6, Q59012-23 and Q57522-3.

SQ Sequence 393 BP; 87 A; 103 C; 124 G; 79 T;

Query Match 80.0%; Score 16; DB 1; Length 393;

Best Local Similarity 65.0%; Pred. No. 5.7;

Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGAYATTCGNCARGA 20

DB 1 GGGGATGACTTACGACAGA 20

RESULT 5

ID Q59022 standard; DNA; 20 BP.

AC Q59022.

DT 12-APR-1994 (first entry)

DE Degenerate primer for p110 cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

OS ss; amplification.

SQ Synthetic.

PN W09321328-A.
 PD 28-OCT-1993.
 PR 13-APR-1993: G00761.
 PR 13-APR-1992: GB-008135.
 PA (LUDWIG) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otau M, Panayotou G,
 PI Parker PJ, Vollmar S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 2; Page 53; 146pp; English.
 CC Two novel cDNAs related to p110 have been cloned. Degenerate
 CC primers were designed to conserved sequences between human p110 and the
 CC related yeast gene VP534. These were used in RT-PCR using mRNA from the
 CC human cell lines MOLT4 and U937. Two novel cDNAs, P1TR-C and P1TR-F,
 CC related to p110, were isolated.
 CC See also 051155-6, 059012-23 and 057522-3.
 SO Sequence 20 BP; 4 A; 2 C; 6 G; 1 T;

Query Match 80.0%; Score 16; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 3.8;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGNAGYATTCGNCARGA 20
 |||||
 DB 1 GGNAGYATTCGNCARGA 20

RESULT 6
 057012
 ID 057012 standard; cDNA to mRNA; 3498 BP.

AC 057012;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 kD catalytic subunit cDNA.
 KM 110 kD catalytic subunit: phosphatidylinositol 3-kinase;
 KM transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KM Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KM blood vessel plaques; ss.
 OS Bos taurus.

FT Key 1. 3307
 FT cds /tag- a
 FT /product- p110

PN W09403609-A.
 PD 17-FEB-1994.
 PR 05-AUG-1993: G01651.
 PR 05-AUG-1992: GB-016654.
 PA (TMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PW, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB; R46294.

PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers

PS Disclosure: Fig 1; 71pp; English.

CC This sequence encodes the 110 kD catalytic subunit of the phosphatidylyl
 CC inositol (Ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.

SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 80.0%; Score 16; DB 1; Length 3498;

Best Local Similarity 65.0%; Pred. No. 7.7;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 OY 1 GGNAGYATTCGNCARGA 20
 |||||
 DB 2410 GGGATGATTCAGCCGCAAGA 2429

RESULT 7
 T58545
 ID T58545 standard; cDNA; 4134 BP.

AC T58545;
 DT 25-MAR-1997 (first entry)
 DE Human phosphatidylinositol 3-kinase PI3K-gamma cDNA.
 KM Phosphatidylinositol 3-kinase; gamma; PI3K; immunogen; immunoassay;
 KM cell proliferation; receptor-mediated signal transduction;
 KM histamine secretion; nerve cell differentiation; glucose transport;
 KM modulation; regulation; Alzheimer's disease; lipolysis; ds.
 OS Homo sapiens.

FT Key Location/Qualifiers
 FT cds 423. 3572
 FT /tag- a
 FT /product- PI3K-gamma

PN DE4445562-C1.
 PD 04-APR-1996.
 PR 20-DEC-1994: 445562.
 PR 13-OCT-1994: DE-436696.
 PA (PLAC) MAX PLANCK GES FORDERUNG WISSENSCHAFTEN.
 PI Hanck T, Stoyanov B, Wetzel R;
 DR WPI: 96-172545/18.
 DR P-PSDB; W11576.

PT New phosphatidylinositol 3-kinase protein - useful as immunogen and
 PT for determ. of kinase activity

PS Claim 5; Page 7; 10pp; German.

CC A 402 bp cDNA fragment was amplified from a human bone marrow
 CC library using PCR primers corresponding to amino acid sequences
 CC KNGDDR and HIDEG. The amplified fragment was used to probe a human
 CC cDNA library and several overlapping clones were isolated.
 CC The largest clone had the present sequence and coded for a protein
 CC of 1049 residues. The protein is a novel phosphatidylinositol 3-
 CC kinase (PI3K) that differs in its regulatory mechanism from the
 CC known PI3K-alpha and -beta enzymes. The new enzyme has been
 CC designated PI3K-gamma and can be used as an immunogen. The enzyme,
 CC antibodies against it or nucleic acid encoding it can be used for
 CC modulating cell proliferation, receptor-mediated signal transduction,
 CC histamine secretion, nerve cell differentiation, glucose transport
 CC and anti-lipolytic activity or for treating Alzheimer's disease.
 CC N.B. Although the claimed sequences are referred to by SEQ. ID.
 CC numbers, a sequence listing did not appear in the original printed
 CC patent application.
 SO Sequence 4134 BP; 1127 A; 999 C; 998 G; 1010 T;

Query Match 80.0%; Score 16; DB 1; Length 4134;
 Best Local Similarity 65.0%; Pred. No. 7.8;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 GGNAGYATTCGNCARGA 20.
 |||||
 DB 2620 GGTATGATTCGCCCAAGA 2639

RESULT 8
 T58546

ID T58546 standard; cDNA; 4137 BP.

AC T58546;
 DT 25-MAR-1997 (first entry)
 DE Human phosphatidylinositol 3-kinase PI3K-gamma cDNA.
 KM Phosphatidylinositol 3-kinase; gamma; PI3K; immunogen; immunoassay;
 KM cell proliferation; receptor-mediated signal transduction;
 KM histamine secretion; nerve cell differentiation; glucose transport;
 KM modulation; regulation; Alzheimer's disease; lipolysis; ds.
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT cds 423..3575
 FT /*tag= a
 FT /product= PI3K-gamma
 PN DE4445562-C1.
 PD 04-APR-1996.
 PR 20-DEC-1994: 445562.
 PR 13-OCT-1994: DE-436696.
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PI Hancock T, Stoyanov B, Wetzker R;
 DR WPI: 96-172545/18.
 DR P-PSDB: W11577.
 PT New phosphatidylinositol 3-kinase protein - useful as immunogen and for dephosphorylation of kinase activity
 PS Claim 5; Page -: 10pp; German.
 CC A 402 bp cDNA fragment was amplified from a human bone marrow library using PCR primers corresponding to amino acid sequences CC KNGDDR and HIRDFG. The amplified fragment was used to probe a human CC 0937 cell cDNA library and several overlapping clones were isolated. CC The largest clone coded for a protein of 1049 residues. The protein CC is a novel phosphatidylinositol 3-kinase (PI3K) that differs in its CC regulatory mechanism from the known PI3K-alpha and -beta enzymes. CC The new enzyme has been designated PI3K-gamma. Another clone, coding CC for a PI3K-gamma of 1050 residues and having the present CC sequence, was then isolated. The enzyme can be used as an immunogen. CC The enzyme, antibodies against it or nucleic acid encoding it can be CC used for modulating cell proliferation, receptor-mediated signal CC transmission, histamine secretion, nerve cell differentiation, CC glucose transport and anti-lipolytic activity or for treating CC Alzheimer's disease.
 CC N.B. Although the claimed sequences are referred to by SEQ.ID.
 CC Numbers, a sequence listing did not appear in the original printed CC patent application.
 SQ Sequence 4137 BP; 1127 A; 999 C; 1001 G; 1010 T;

Query Match 80.0%; Score 16; DB 1; Length 4137;
 Best Local Similarity 65.0%; Pred. No. 7.8;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGYATYTRCGNCARGA 20
 DB 2823 GGTGATGATCTGCCCANAGA 2842

RESULT 9
 V04634
 ID V04634 standard; cDNA; 3808 BP.
 AC V04634:
 DT 17-AUG-1998 (first entry)
 DE Porcine phosphoinositide 3OH-kinase p120 subunit cDNA.
 KW G-beta-gamma regulated phosphatidylinositol-3' kinase; pig;
 KW phosphoinositide 3OH-kinase; PI3K; signal transduction;
 KW phosphatidylinositol (3,4,5)-trisphosphate; G-protein; receptor;
 KW transgenic animal; knockout animal; inflammation; arthritis;
 KW septic shock; adult respiratory distress syndrome; pneumonia;
 KW asthma; allergy; reperfusion injury; atherosclerosis; cancer;
 KW Alzheimer's disease; cancer; antisense; ribozyme; diagnosis;
 KW therapy; drug screening; ss.
 OS Sus scrofa.
 FH Key Location/Qualifiers
 FT CDS 89..3397
 FT /*tag= a
 FT W09749618-A2.
 PD 31-DEC-1997.
 PR 26-JUN-1997: U11219.
 PR 27-JUN-1996: US-672211.
 PA (ONXX-) ONXX PHARM.
 PI Bräselemann S, Hawkins PT, Stephens L;
 DR WPI: 98-077181/07.
 DR P-PSDB: W33948.
 PT DNA encoding G-beta-gamma regulated phosphatidylinositol-3' kinase, p101 and p120 subunits - useful for diagnosis, drug screening,

PT clinical trial monitoring and treatment of inflammatory disorders
 PS Claim 54; Fig 3A-C; 151pp; English.
 CC This cDNA sequence codes for the p120 catalytic subunit (see CC W23948) of pig G-protein regulated phosphatidylinositol-3' kinase, CC a heterodimeric enzyme which produces the intracellular messenger CC phosphatidylinositol (3,4,5)-trisphosphate in response to activation CC of trimeric G protein-linked receptors. This novel protein, which CC also contains a regulatory subunit, p101 (see V04633), is found in CC cells of hematopoietic origin and is involved in immune system CC responses which cause inflammation. p120 cDNA was obtained from a CC pig neutrophil cDNA library using a degenerate oligonucleotide CC probe (see T99713) based on an isolated p120 tryptic peptide. The CC p120 cDNA clone in pCMV3mycP120 is deposited as ATCC 97637. The CC invention encompasses pig and human p101 and p120 nucleotides, CC host cell expression systems, p101 and p120 proteins (see CC W23946-49), fusion proteins, polypeptides and peptides, antibodies CC to these proteins, and transgenic animals and knockout animals. CC Compounds which are useful for treating inflammatory response CC disorders can be identified by screening assays using a G protein CC activated PI3K, or a cultured host cell that expresses the p101 CC gene. Antagonists of G protein stimulated PI3K (acting through the CC p101 subunit, especially by disrupting the interaction between the CC p101 and p120 subunits) can be used to treat arthritis, septic CC shock, adult respiratory distress syndrome (ARDS), pneumonia, CC asthma, allergies, reperfusion injury, atherosclerosis, cancer and CC Alzheimer's disease. The nucleic acids and their products can also CC be used for diagnosis, drug screening and clinical trial monitoring CC of inflammatory diseases.
 SQ Sequence 3808 BP; 1010 A; 965 C; 953 G; 880 T;

Query Match 80.0%; Score 16; DB 1; Length 3808;
 Best Local Similarity 65.0%; Pred. No. 7.7;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNAGYATYTRCGNCARGA 20
 DB 2591 GGTGATGATCTGCCCANAGA 2610

RESULT 10
 V74100
 ID V74100 standard; DNA; 3808 BP.
 AC V74100:
 DT 09-MAR-1999 (first entry)
 DE Porcine G-protein regulated PI3K p120 adapter subunit DNA.
 KW Phosphoinositide 3-hydroxylase; PI3K; trimeric G protein; porcine;
 KW adapter subunit; regulatory subunit; p101; p120; catalytic subunit;
 KW detection; diagnosis; activation disorder; hematopoietic system;
 KW treatment; immune disorder; inflammation; arthritis; septic shock;
 KW adult respiratory distress syndrome; pneumonia; asthma; allergy;
 KW reperfusion injury; atherosclerosis; Alzheimer's disease; cancer; ss.
 OS Sus scrofa.
 PN US5856132-A.
 PD 05-JAN-1999.
 PR 15-AUG-1997; 916917.
 PR 15-AUG-1997; US-916917.
 PR 27-JUN-1996; US-672211.
 PA (ONXX-) ONXX PHARM.
 PI Bräselemann S, Hawkins PT, Stephens L;
 DR WPI: 99-105107/09.
 DR P-PSDB: W90083.
 PT Nucleic acid encoding regulatory (p101) and catalytic (p120) PT subunits of a heterodimeric phosphatidylinositol-3' kinase - useful PT in treatment and diagnosis of immune system disorders, e.g. PT arthritis, cancer and Alzheimer's disease
 PS Example IX; Fig 3A-C; 75pp; English.
 CC This sequence encodes a novel catalytic subunit, p120, from porcine CC phosphoinositide 3-hydroxylase (PI3K) which is regulated by beta-gamma CC subunits of trimeric G-protein. Nucleic acid coding for p101 or p120, or CC their fragments, are used as probes and primers for identifying p101 or CC p120 gene mutations, allelic variations or regulatory defects, CC particularly for the diagnosis of activation disorders (or

CC susceptibility) in cells of the haematopoietic system. The related
 CC proteins, antibodies, agonists and antagonists can be used similarly. The
 CC p101 and p120 proteins, peptides or fusion proteins are used to treat or
 CC screen for potential agents for treating immune disorders, particularly
 CC inflammation, e.g. arthritis, septic shock, adult respiratory distress
 CC syndrome, pneumonia, asthma, allergies, reperfusion injury,
 CC atherosclerosis, Alzheimer's disease and cancer.
 SQ Sequence 3808 BP; 1010 A; 965 C; 953 G; 880 T;

Query Match 80.0%; Score 16; DB 1; Length 3808;
 Best Local Similarity 65.0%; Pred. No. 7.7;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20
 |||:||||:|||||
 Db 2591 GGTGACGATCTCGCCAGCA 2610

RESULT 11
 V74104
 ID V74104 standard; DNA: 5162 BP.

AC V74104:
 DT 09-MAR-1999 (first entry)
 DE Human G-protein regulated PI3K p120 adapter subunit DNA.
 KW Phosphoinositide 3-hydroxylase; PI3K; trimeric G protein; human;
 KW adapter subunit; regulatory subunit; p101; p120; catalytic subunit;
 KW detection; diagnosis; activation disorder; haematopoietic system;
 KW treatment; immune disorder; inflammation; arthritis; septic shock;
 KW adult respiratory distress syndrome; pneumonia; asthma; allergy;
 KW reperfusion injury; atherosclerosis; Alzheimer's disease; cancer; ss.
 OS Homo sapiens.
 PN US856132-A.
 PD 05-JAN-1999.
 PE 15-AUG-1997; 916917.
 PR 15-AUG-1997; US-916917.
 PR 27-JUN-1996; US-672211.
 PA (ONTX-) ONTX PHARM.
 PI Braseleiman S, Hawkins PT, Stephens L;
 DR WPI: 99-105107/09.
 DR P-PSDB: M90089.
 PT Nucleic acid encoding regulatory (p101) and catalytic (p120)
 PT subunits of a heterodimeric phosphatidylinositol-3' kinase - useful
 PT in treatment and diagnosis of immune system disorders, e.g.
 PT arthritis, cancer and Alzheimer's disease
 PS Example X: Fig 12A-C; 75pp; English.
 CC This sequence encodes a novel catalytic subunit, p120, from human
 CC phosphoinositide 3-hydroxylase (PI3K) which is regulated by beta-gamma
 CC subunits of trimeric G-protein. Nucleic acid coding for p101 or p120, or
 CC their fragments, are used as probes and primers for identifying p101 or
 CC p120 gene mutations, allelic variations or regulatory defects,
 CC particularly for the diagnosis of activation disorders (or
 CC susceptibility) in cells of the haematopoietic system. The related
 CC proteins, antibodies, agonists and antagonists can be used similarly. The
 CC p101 and p120 proteins, peptides or fusion proteins are used to treat or
 CC screen for potential agents for treating immune disorders, particularly
 CC inflammation, e.g. arthritis, septic shock, adult respiratory distress
 CC syndrome, pneumonia, asthma, allergies, reperfusion injury,
 CC atherosclerosis, Alzheimer's disease and cancer.
 SQ Sequence 5162 BP; 1471 A; 1173 C; 1149 G; 1367 T;

Query Match 80.0%; Score 16; DB 1; Length 5162;
 Best Local Similarity 65.0%; Pred. No. 8.1;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20
 |||:||||:|||||
 Db 2586 GGTGATGATCTCGCCAGCA 2605

RESULT 12
 060936

ID 060936 standard; DNA: 423 BP.

AC 060936:
 DT 16-MAR-1994 (first entry)
 DE Human brain Expressed Sequence Tag EST01019.
 KW Gene transcription product; genetic markers; tagging; in vivo;
 KW transcription; mapping; locations; chromosomes; chromosomal; ss.
 OS Homo sapiens.
 PN W09316178-A.
 PD 19-AUG-1993.
 PE 12-FEB-1993; U01294.
 PR 12-FEB-1992; US-837195.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICE.
 PI Adams MD, Moreno RF, Venter CJ;
 DR WPI: 93-272882/34.
 PT Enriched oligonucleotides and corresp. sequences - used as
 PT markers for human genes transcribed in-vivo, facilitate tagging
 PT of most human genes
 PS Example 4: Page 413; 500pp; English.
 CC The Expressed Sequence Tag was isolated from a human brain cDNA
 CC library as part of a large set of ESTs which can be used as markers
 CC for human genes transcribed in vivo. They can be used to facilitate
 CC tagging of most human genes, for mapping locations of expressed genes
 CC on chromosomes, for individual or forensic identification, for mapping
 CC locations of disease-associated genes, for identification of tissue
 CC type, and for prep. of antisense sequences, probes and constructs.
 CC EST01019 has an "excellent" coding probability as evaluated using the
 CC coding-region prediction program CRM. See also Q59041-Q61440.
 SQ Sequence 423 BP; 87 A; 122 C; 124 G; 89 T;

Query Match 74.0%; Score 14.8; DB 1; Length 423;
 Best Local Similarity 65.0%; Pred. No. 27;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20
 |||:||||:|||||
 Db 272 GGAGACGATCTCGCCAGCA 291

RESULT 13

ID T96605/c
 AC T96605 standard; DNA: 1612 BP.
 DT 19-FEB-1998 (first entry)
 DE C. glutamicum atrop gene.
 KW Corynebacterium glutamicum; atrop gene; transporter; aromatic amino acid;
 KW microbial amino acid production; secretion rate; tryptophan; tyrosine;
 KW phenylalanine; feed additive; aspartame; ss.
 OS Corynebacterium glutamicum ATCC 13032.
 FH Key Location/Qualifiers
 FT CDS 80..1471
 FT /*tag= a
 PM DE19523279-A1.
 PD 09-JAN-1997.
 PE 27-JUN-1995; 023279.
 PR 27-JUN-1995; DE-023279.
 PA (KERU) FORSCHUNGSZENTRUM JUELICH GMBH.
 PI Eggeling L, Kraemer R, Wehrmann A;
 DR WPI: 97-066157/07.
 DR P-PSDB: W36452.
 PT Microbial prodn. of (aromatic) amino acids in cell transformed with
 PT gene for transport protein - provides increased secretion of product
 PT into culture medium
 PS Claim 8: Page 5; 7pp; German.
 CC The present sequence contains the Corynebacterium glutamicum ATCC 13032
 CC atrop gene which encodes a transporter protein for aromatic amino acid.
 CC This gene is useful in the microbial production of amino acids. The
 CC expression of the gene increases the secretion rate of the amino acid
 CC and hence the cell excretes a greater amount of the amino acid into the
 CC medium. In particular microbial production of e.g. tryptophan (for
 CC pharmaceuticals or feed additives), tyrosine (for pharmaceuticals or
 CC intermediates) or phenylalanine (an intermediate for aspartame), is
 CC preferred.

SO Sequence 1612 BP; 280 A; 450 C; 421 G; 461 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1612;

Best Local Similarity 65.0%; Pred. No. 32;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGAGYGYTTCGNCARGA 20

DB 235 GGGAGTGTACGCCAGCA 216

RESULT 14

VO1877

VO1877 standard; DNA; 3602 BP.

DT 20-APR-1998 (first entry)

DE Human NP1K gene.

KM Human; foetal brain cDNA library; GDP dissociation stimulating protein;

KW brain specific nucleosome assembly protein; diagnosis; therapy;

KM skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NP1K;

KW nel-related protein type 1; nel-related type 2; hereditary disease;

AW cancer; ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 429..2882

FT /tag- a

FT /product- "NP1K"

EP-796913-A2.

PD 24-SEP-1997.

PF 19-MAR-1997; 104842.

PR 05-MAR-1997; JP-069163.

PR 19-MAR-1996; JP-063410.

PA (SARA) OTSUKA PHARM CO LTD.

PI Fujiwara T, Horie M, Watanabe T;

DR WPI: 97-459830/43.

P-PSDB; W37498.

PT Novel human genes, e.g. brain-specific nucleosome assembly protein -

PT useful for diagnosis or therapy of hereditary disease and cancer

PS Claim 15; Page 78-83; 123pp; English.

CC The present sequence encodes a NP1K isolated from a human foetal brain

CC cDNA library. The nucleotide or amino acid sequences are useful for

CC in-vitro diagnosis of hereditary diseases and cancer and for preparation

CC of pharmaceuticals.

Sequence 3602 BP; 849 A; 1006 C; 985 G; 762 T;

Query Match 74.0%; Score 14.8; DB 1; Length 3602;

Best Local Similarity 65.0%; Pred. No. 36;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGAGYGYTTCGNCARGA 20

DB 2124 GGGAGTGTACGCCAGCA 2143

RESULT 15

VO1878

VO1878 standard; cDNA; 2487 BP.

AC VO1878;

DT 20-APR-1998 (first entry)

DE Human NP1K gene.

KM Human; foetal brain cDNA library; GDP dissociation stimulating protein;

KW brain specific nucleosome assembly protein; diagnosis; therapy;

KM skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NP1K;

KW nel-related protein type 1; nel-related type 2; hereditary disease;

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 1..2487

FT /tag- a

FT /product- "NP1K"

FT /note- "no stop codon given"

PN EP-796913-A2.

PD 24-SEP-1997.

PF 19-MAR-1997; 104842.

PR 05-MAR-1997; JP-069163.

PR 19-MAR-1996; JP-063410.

PA (SARA) OTSUKA PHARM CO LTD.

PI Fujiwara T, Horie M, Watanabe T;

DR WPI: 97-459830/43.

P-PSDB; W37499.

PT Novel human genes, e.g. brain-specific nucleosome assembly protein -

PT useful for diagnosis or therapy of hereditary disease and cancer

PS Claim 17; Page 87-88; 123pp; English.

CC The present sequence encodes a NP1K isolated from a human foetal brain

CC cDNA library. The nucleotide or amino acid sequences are useful for

CC in-vitro diagnosis of hereditary diseases and cancer and for preparation

CC of pharmaceuticals.

Sequence 2487 BP; 604 A; 682 C; 664 G; 537 T;

Query Match 74.0%; Score 14.8; DB 1; Length 2487;

Best Local Similarity 65.0%; Pred. No. 34;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGAGYGYTTCGNCARGA 20

DB 1732 GGGAGTGTACGCCAGCA 1751

Search completed: September 13, 1999, 15:59:37
Job time: 4804 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:57 ; Search time 1694.61 Seconds

(without alignments)
23.280 Million cell updates/sec

Title: US-09-325-095-27

Perfect score: 20

Sequence: 1 GGNAGYATTCGNCARCA 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: em_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	16	80.0	324	20	T29663	T29663 EST89481 Hu
2	16	80.0	424	22	H14711	H14711 YM24A02.r1
3	16	80.0	375	28	C12533	C12533 C12533 Yu1
4	16	80.0	667	30	AA206574	AA206574 zq51902.r
5	16	80.0	597	30	AA210248	AA210248 mc84c09.r
6	16	80.0	437	35	AA542224	AA542224 VK28A04.r
7	16	80.0	357	35	AA571816	AA571816 V191F09.r
8	16	80.0	574	37	AB009107	AB009107 AB009107
9	16	80.0	413	39	AA661836	AA661836 ak39A08.S
10	16	80.0	497	47	AI531818	AI531818 SC03087.S
11	14.8	74.0	423	20	M78871	M78871 EST01019 Su
12	14.8	74.0	271	20	Z42324	Z42324 HSCOF081 n
13	14.8	74.0	223	23	D57845	D57845 HUM329H07B
14	14.8	74.0	396	24	D61665	D61665 HUM428B12B
15	14.8	74.0	413	30	AA231198	AA231198 mw38f12.r
16	14.8	74.0	475	33	AA437822	AA437822 ve34g12.r
17	14.8	74.0	727	41	AI1001270	AI1001270 MEST5-D1.
18	14.8	74.0	750	42	AI158022	AI158022 pd26A08.r
19	14.8	74.0	443	46	AI426863	AI426863 mm97e11.x
20	14.8	74.0	538	51	AI714581	AI714581 605059G11
21	14.4	72.0	384	21	R17530	R17530 yg03d11.r1
22	14.4	72.0	593	24	N24991	N24991 yk16e11.r1
23	14.4	72.0	419	26	W95377	W95377 ze01g12.r1
24	14.4	72.0	451	27	AA001396	AA001396 ze46A05.r
25	14.4	72.0	280	27	AA017578	AA017578 ze37F04.r
26	14.4	72.0	231	31	AA329679	AA329679 EST33318
27	14.4	72.0	441	33	AA452699	AA452699 zX39D09.r
28	14.4	72.0	334	34	AA485157	AA485157 aa40A05.r
29	14.4	72.0	599	38	AA801073	AA801073 EST190570
30	14.4	72.0	302	39	AA842951	AA842951 T4689 MVA
31	14.4	72.0	206	40	AT000268	AT000268 AT000268
32	14.4	72.0	95	44	AI299931	AI299931 cm59P03.x
33	14	70.0	507	37	AA698421	AA698421 HU04A64.5
34	14	70.0	784	41	AI069387	AI069387 mg2e006d
35	14	70.0	349	49	AU055795	AU055795 AU055795
36	14	70.0	167	49	AV017180	AV017180 AV017180
37	13.8	69.0	103	20	D12085	D12085 HUM0516A08
38	13.8	69.0	764	30	AA264673	AA264673 LP08378.5
39	13.8	69.0	512	31	AA310586	AA310586 EST181407
40	13.8	69.0	264	33	AA433315	AA433315 TEUF0148
41	13.8	69.0	360	36	C62862	C62862 C62862 Yu1
42	13.8	69.0	360	36	C64513	C64513 C64513 Yu1
43	13.8	69.0	375	36	C70607	C70607 C70607 Yu1
44	13.8	69.0	362	42	AI145747	AI145747 UT-R-BT0
45	13.6	68.0	424	25	W13743	W13743 mb32c10.r1

ALIGNMENTS

RESULT 1

T29663 LOCUS T29663 324 bp mRNA EST 06-SEP-1995

DEFINITION EST89481 Human Small Intestine Homo sapiens cDNA 5' end similar to phosphatidylinositol 3-kinase p110, beta isoform (HT3411), mRNA

ACCESSION T29663

NTD 9611761

VERSION T29663.1 GI:611761
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 324)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
 Ball,C.J., Lee,N., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
 White,O., Sutton,G., Blake,J.A., Brandon,R.C., Chiu,M.W.,
 Clayton,R.A., Cline,R.T., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
 Fitzgerald,L.M., Fitzhugh,M.M., Fritchman,J.L., Geoghegan,N.S.M.,
 Glodetz,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkley,P.S.,
 Kelley,J.M., Klimek,K.M., Kelley,J.C., Liu,L.T., Marrao,S.M.,
 Merrick,J.M., Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T.,
 Pellegrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L.,
 Seudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R.,
 Weidman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepeda,M.A.,
 Coleman,T.A., Collins,E.-J., Dinko,D., Feng,P., Fertle,A.,
 Fischer,C., Hastings,G.A., He,W.-W., Hu,J.-S., Greene,D.M.,
 Gruber,J., Hudson,P., Kim,A., Kozak,D.L., Kunsch,C., Li,H., Li,H.,
 Weissner,P.S., Olsen,H., Raymond,L., Wei,Y.-F., Wang,J., Xu,C.,
 Yu,G.-L., Ruben,S.M., Dillon,P.J., Fannon,M.R., Rosen,C.A.,
 Hasselting,W.A., Fields,C., Fraser,C.M. and Venter,J.C.
 Initial Assessment of Human Gene Diversity and Expression Patterns
 Based Upon 83 Million Basepairs of cDNA Sequence
 Nature 377, 3-174 (1995)
 JOURNAL MEDLINE
 COMMENT 96026280
 TITLE Contact: Venter, JC
 The Institute for Genomic Research
 932 Clopper Rd, Gaithersburg, MD 20878
 Tel: 3018690956
 Fax: 3018699423
 Email: tdbinfo@tdb.tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please contact the TIGR Database
 (tdbinfo@tdb.tigr.org)
 Seq primer: M13 Reverse.
 Location/Qualifiers
 source
 1.324
 /organism="Homo sapiens"
 /db_xref="ATCC (lnhost):107054"
 /db_xref="taxon:9606"
 /clone_lib="Human Small Intestine"
 /note="Organ: small intestine"
 /note="Organ: small intestine" 80 g 96 t 2 others
 BASE COUNT 88 a 58 c 80 g 96 t
 ORIGIN
 Query Match 80.0%; Score 16; DB 20; Length 324;
 Best Local Similarity 65.0%; Pred. No. 43;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGNAGYAGYTRCGNCARGA 20
 |||:||||:|||||
 Db 70 GGTGATGATTACGACAGCA 89
 RESULT 2
 LOCUS H14711 424 bp mRNA EST 27-JUN-1995
 DEFINITION YMA24602.f1 Soares Infant brain INIB Homo sapiens cDNA clone
 IMAGE:48772 5' similar to SP:P100_BOVIN P22871 PHOSPHATIDYLINOSITOL
 3-KINASE CATALYTIC SUBUNIT ;, mRNA sequence.
 ACCESSION H14711
 NID 9879531
 VERSION H14711.1 GI:879531
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 424)
 Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
 Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and
 Wilson,R.
 The WashU-Merck EST Project
 Unpublished (1993)
 On Sep 21, 1992 this sequence version replaced gi:276288.
 JOURNAL COMMENT
 TITLE Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Insert Size: 1696
 High quality sequence stops: 312
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1696 Std Error: 0.00
 Seq primer: M13Rp1
 High quality sequence stop: 312.
 Location/Qualifiers
 source
 1.424
 /organism="Homo sapiens"
 /db_xref="GDB:421313"
 /db_xref="taxon:9606"
 /clone="IMAGE:48772"
 /clone_lib="Soares Infant brain INIB"
 /sex="female"
 /dev_stage="73 days post natal"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: whole brain; Vector: latmid BA; Site: 1: Not
 I; Site 2: Hind III; 1st strand cDNA was primed with a Not
 I - oligo(dT) primer 15',
 AACTGGAAGATTCGCGCGCGCAGCAATTTTCTTTTCTTTT 3'];
 double-stranded cDNA was ligated to Hind III adaptors
 (pharmacia), digested with Not I and directionally cloned
 into the Not I and Hind III sites of the latmid BA vector.
 Library went through one round of normalization. Library
 constructed by Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 107 a 85 c 104 g 124 t 4 others
 ORIGIN
 Query Match 80.0%; Score 16; DB 22; Length 424;
 Best Local Similarity 65.0%; Pred. No. 47;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GGNAGYAGYTRCGNCARGA 20
 |||:||||:|||||
 Db 198 GGTGATGATTACGACAGCA 217
 RESULT 3
 LOCUS C12533 375 bp mRNA EST 28-DEC-1998
 DEFINITION C12533 Yui1 Kohara unpublished cDNA Caenorhabditis elegans cDNA
 clone yk152b12 5', mRNA sequence.
 ACCESSION C12533
 NID 91560086
 VERSION C12533.1 GI:1560086
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitidae;
 Rhabdilitina; Rhabdilitidae; Rhabdilitidae; Peloderidae; Caenorhabditis.
 Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
 Sano,M., Miyata,A. and Nishigaki,A.
 Expression map of the C.elegans genome

JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1288981.

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.jp.

FEATURES

source
 1. 375
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him 8(elt489)"
 /db_xref="taxon:6239"
 /clone_lib="Yuji Kohara unpublished cDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male,
 tissue_type=whole animal"
 BASE COUNT 114 a 61 c 79 g 115 t 6 others
 ORIGIN

Query Match 80.0%; Score 16; DB 28; Length 375;
 Best Local Similarity 65.0%; Pred. No. 45;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 GGNAGYATTCGNCARCA 20
 |||:::|||||
 Db 11 GGAGATGATTACGCACAGA 30

RESULT 4
 LOCUS AA206574 667 bp mRNA EST 12-MAR-1998
 DEFINITION 2951902.r1 Striatum neuroepithelium (#937231) Homo sapiens CDNA
 clone IMAGE:645170 5' similar to SW:P11B_HUMAN P42338
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, BETA ISOFORM ;
 mRNA sequence.

ACCESSION AA206574
 NID 91801954
 VERSION AA206574.1 GI:1801954
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Getsel,G., Jost,S.,
 Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
 Martin,J., Moore,B., Scheinberg,R., Steptoe,M., Tan,F.,
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 * Unpublished (1997)
 On Apr 14, 1993 this sequence version replaced gi:692719.

COMMENT
 TITLE JOURNAL
 CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewartson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 2550 Std Error: 0.00
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 297.
 Location/Qualifiers

FEATURES

source
 1. 667
 /organism="Homo sapiens"
 /db_xref="GDB:5215421"
 /db_xref="taxon:9606"
 /map="688H07; 2"

/clone="IMAGE:645170"
 /clone_lib="Striatum neuroepithelium (#937231)"
 /dev_stage="Neura-2/RA neuroepithelial cells"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site1: EcoRI; Site2:
 XhoI; Cloned unidirectionally. Primer: Oligo dT NT2
 cells (Neura-2/c1.D1) induced with Retinoic acid for 24
 hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector; -5'
 adaptor sequence: 5' GAATTCGACGACG 3' -3' adaptor
 sequence: 5' CTCGACTTTTCTTTTCTTTT 3'"
 BASE COUNT 176 a 139 c 162 g 179 t 11 others
 ORIGIN

Query Match 80.0%; Score 16; DB 30; Length 667;
 Best Local Similarity 65.0%; Pred. No. 54;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;
 Oy 1 GGNAGYATTCGNCARCA 20
 |||:::|||||
 Db 213 GGATGATTTACGCACAGA 232

RESULT 5
 LOCUS AA210248 597 bp mRNA EST 29-JAN-1997
 DEFINITION m084c09.r1 Beddington mouse embryonic region Mus musculus CDNA
 clone IMAGE:560272 5' similar to TR:G987948 G987948
 PHOSPHATIDYLINOSITOL 3-KINASE. ; mRNA sequence.

ACCESSION AA210248
 NID 91807546
 VERSION AA210248.1 GI:1807546
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Getsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Scheinberg,R., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 The WashU-HMI Mouse EST Project
 * Unpublished (1996)
 On Sep 12, 1996 this sequence version replaced gi:1394729.

COMMENT
 TITLE JOURNAL
 CONTACT: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:341064
 Possible reversed clone: similarity on wrong strand
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 338.
 Location/Qualifiers

FEATURES

source
 1. 597
 /organism="Mus musculus"
 /strain="C57BL6 x DBA"
 /db_xref="taxon:10090"
 /clone="IMAGE:560272"
 /clone_lib="Beddington mouse embryonic region"
 /sex="pooled"
 /tissue_type="embryo"
 /dev_stage="7.5dpc"
 /lab_host="DH12S"
 /note="Organ: whole embryo; Vector: PCMV-SPORT; Site1:
 SalI; Site2: NotI; Cloned unidirectionally. Primer:

Query Match 80.0%; Score 16; DB 35; Length 357;
 Best Local Similarity 65.0%; Pred. No. 44;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20
 |||||:||||:||||
 Db 40 GGAGACGACTTGGCTCAGCA 59

RESULT 8
 AB009107 574 bp mRNA EST 05-DEC-1997
 LOCUS AB009107 Chlamydomonas W80 lambda ZAP II Chlamydomonas sp. cDNA
 DEFINITION similar to phosphatidylinositol 3-kinase, mRNA sequence.
 ACCESSION AB009107
 MID 92662210
 VERSION AB009107.1 GI:2662210
 KEYWORDS EST.
 SOURCE Chlamydomonas sp.
 ORGANISM Chlamydomonas sp.
 Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 Chlamydomonadaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 574)
 *Kanaboshi, H., Ikeda, K. and Miyasaka, H.
 Isolation of several anti-stress genes from the halotolerant green
 alga Chlamydomonas by a simple functional expression screening in
 E. coli
 Unpublished (1998)
 JOURNAL On Sep 19, 1997 this sequence version replaced gi:1520267.

COMMENTS

FEATURES
 source
 1..574
 /organism="Chlamydomonas sp."
 /strain="W80"
 /db_xref="taxon:3056"
 /clone_11b="Chlamydomonas W80 lambda ZAP II"
 /note="Vector: lambda ZAP II; The cDNA clone was isolated
 from the halotolerant green alga Chlamydomonas W80 by a
 functional expression screening in E. coli cells. The
 principle of the screening method was based on the
 acquirement of stress tolerance of the bacterial cells
 carrying the cDNA."

BASE COUNT 96 a 197 c 184 g 97 t

ORIGIN

Query Match 80.0%; Score 16; DB 37; Length 574;
 Best Local Similarity 65.0%; Pred. No. 51;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20
 |||||:||||:||||
 Db 110 GGCGACGACCTGGCGACAGA 129

RESULT 9
 AA861836 413 bp mRNA EST 04-JAN-1999
 LOCUS AA861836 si Soares-testis_NHT Homo sapiens cDNA clone IMAGE:1408310
 DEFINITION 3' similar to SW-PIIB_HUMAN P42338 PHOSPHATIDYLINOSITOL 3-KINASE
 CATALYTIC SUBUNIT, BETA ISOFORM ;contains Alu repetitive element;;
 mRNA sequence.
 ACCESSION AA861836
 MID 92953976
 VERSION AA861836.1 GI:2953976
 KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 413)
 *NCI-CCAP http://www.ncbi.nlm.nih.gov/nciccap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397601.

CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bhrp/image/image.html

FEATURES
 source
 1..413
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_11b="Soares-testis_NHT"
 /sex="male"
 /lab_host="DH10B"
 /note="Vector: pRT3D-Pac (Pharmacia) with a modified
 polylinker. Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from Clontech
 Laboratories, Inc., and primed with a Not I - oligo(dT)
 primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pRT3D vector. Library
 went through one round of normalization to Cots, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 116 a 93 c 74 g 130 t

ORIGIN

Query Match 80.0%; Score 16; DB 39; Length 413;
 Best Local Similarity 65.0%; Pred. No. 46;
 Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTRCGNCARGA 20
 |||||:||||:||||
 Db 251 GGTGATGATTCGACAGCA 232

RESULT 10
 A1531818 497 bp mRNA EST 18-MAR-1999
 LOCUS A1531818 SD03087.5prine SD Drosophila melanogaster Schneider L2 cell culture
 DEFINITION P072 Drosophila melanogaster cDNA clone SD03087 5prine similar to
 X99912; P13K59F FBgn0015277 P1D:258266 SPTRXBL:P91635, mRNA
 sequence.
 ACCESSION A1531818
 MID 94445953
 VERSION A1531818.1 GI:4445953
 KEYWORDS EST.
 SOURCE fruit fly.
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

REFERENCE Muscomorpha: Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 497)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Bokstein, P., Lewis, S., and Rubin, G. M.
TITLE BDGP/HMI Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT On Mar 10, 1998 this sequence version replaced g1:2948480.

CONTACT: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST_estfruitfly.berkeley.edu
hit genomic sequence AC005639
Plate: 30 row: H column: 3
High quality sequence stop: 476.
location/Qualifiers
1.497
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="SD03087"
/clone_1lb="SD Drosophila melanogaster Schneider L2 cell
culture P0T2"
/lab_host="DH5-alpha"
/note="Vector: P0T2; Site-1: EcoRI; Site-2: XhoI; Sized
fractionated cDNAs were directly ligated into P0T2.
Plasmid cDNA library."
QY 1 GGNAGAGAYTTCGNCARGA 20
11 |||:||||:|||||
Db 371 GGATGATGATCGCCGACGA 390

Query Match 80.0%; Score 16; DB 47; Length 497;
Best Local Similarity 65.0%; Pred. No. 49;
Matches 13; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

RESULT 11
LOCUS M78871 423 bp mRNA EST 26-MAY-1992
DEFINITION EST01019 Subtracted Hippocampus, Strata gene (cat. #936205) Homo
sapiens cDNA clone HHCPA26, mRNA sequence.
ACCESSION M78871
NID 9273185
VERSION M78871.1 GI:273185
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 423)
AUTHORS Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M.,
Ullrich, B., Negle, J.M., Fields, C., and Venter, J.C.
TITLE Sequence identification of 2,375 human brain genes
JOURNAL Nature 355, 633-634 (1992)
MEDLINE 92168112
COMMENT Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
Seq primer: M13 forward.
Location/Qualifiers
1.423
/organism="Homo sapiens"
/db_xref="ATCC (inhost):78337"

/db_xref="GDB:D0S1850E"
/db_xref="taxon:9606"
/clone="HHCPA26"
/clone_1lb="Subtracted Hippocampus, Strata gene (cat.
#936205)"
/note="Vector: lambdaZAP-II; The hippocampus library (#4
above) was substracted with a fibroblast cell line cDNA
library (Stratagene cat. #936209; W38 lung fibroblast
cell line; oligo-dt + random primed cDNA synthesis;
lambdaZAP-II vector, 1.0kb average insert size.) by the
method of Sive & St. John (Nucl. Acids Res. 16:10937,
1988)."
BASE COUNT 87 a 122 c 124 g 89 t 1 others
ORIGIN

Query Match 74.0%; Score 14.8; DB 20; Length 423;
Best Local Similarity 65.0%; Pred. No. 2.2e+02;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGNAGAGAYTTCGNCARGA 20
11 |||:||||:|||||
Db 272 GGAGACGACTTCGCGACGA 291
RESULT 12
LOCUS 242324 271 bp mRNA EST 09-NOV-1994
DEFINITION HSCOFF081 normalized infant brain cDNA Homo sapiens cDNA clone
c-0ff08, mRNA sequence.
ACCESSION 242324
NID 9565737
VERSION 242324.1 GI:565737
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 271)
AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B.,
Lorenzo, F., Mitchell, H., Marigault-Samson, R., Pietu, G., Pouliot, Y.,
Sebastien-Kabakichis, C. and Tessier, A.
TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT On Sep 21, 1992 this sequence version replaced g1:279243.

CONTACT: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr
Single read.
Genexpress_1lb: y1c-0ff08
Seq primer: (-21)M13 universal.
Location/Qualifiers
1.271
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-0ff08"
/clone_1lb="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lambda ZAP; Site-1: HindIII;
Site-2: NotI; sex=Female; dev_stage=3 months old;
isolate-muscular atrophy patient; tissue-type-total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the

BASE COUNT 61 a 83 c 75 g 51 t 1 others

ORIGIN

Query Match 74.0%; Score 14.8; DB 20; Length 271;
Best Local Similarity 65.0%; Pred. No. 1.9e+02;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGYATYTCGNCARGA 20
Db 249 GGGGATGACCTTCGGCARGA 268

RESULT 13
LOCUS D57845 223 bp mRNA EST 28-AUG-1995
DEFINITION HUM329H07B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens
ACCESSION D57845
NID 9964467
VERSION D57845.1 GI:964467
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., MaeKawa,H., Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
Unpublished (1995)
On May 9, 1995 this sequence version replaced gi:802631.

TITLE JOURNAL
COMMENT

CONTACT: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035
Insert Length: 1248 Std Error: 0.00
High quality sequence stop: 349.
Location/Qualifiers
1. 223
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19"
/clone="GEN-329H07"
/clone_11b="Clontech human aorta polyA+ mRNA (#6572)"

BASE COUNT 86 a 38 c 50 g 49 t

Query Match 74.0%; Score 14.8; DB 23; Length 223;
Best Local Similarity 65.0%; Pred. No. 1.8e+02;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGYATYTCGNCARGA 20
Db 143 GGGGACGAGTTCGCGCARGA 162

RESULT 14
LOCUS D61665 396 bp mRNA EST 14-DEC-1995
DEFINITION HMM428B12B Clontech human fecal brain polyA+ mRNA (#6535) Homo sapiens
ACCESSION D61665
NID 9971029

VERSION D61665.1 GI:971029
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M., Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H., Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y., MaeKawa,H., Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
Unpublished (1995)
On Apr 14, 1993 this sequence version replaced gi:785805.

TITLE JOURNAL
COMMENT

CONTACT: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035
Insert Length: 1028 Std Error: 0.00
High quality sequence stop: 234.
Location/Qualifiers
1. 396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17 cen-gter: 17q21"
/clone="GEN-428B12"
/clone_11b="Clontech human fetal brain polyA+ mRNA (#6535)"

BASE COUNT 90 a 101 c 107 g 98 t

Query Match 74.0%; Score 14.8; DB 24; Length 396;
Best Local Similarity 65.0%; Pred. No. 2.1e+02;
Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGNAGYATYTCGNCARGA 20
Db 108 GGGGATGACCTTCGGCARGA 127

RESULT 15
LOCUS AA231198 413 bp mRNA EST 26-FEB-1997
DEFINITION mw38f12.T1 Soares mouse 3ME12.5 Mus musculus cDNA clone IMAGE:673007 5' similar to gb:X62534 HIGH MOBILITY GROUP PROTEIN HMG2 (HUMAN); gb:246757 M.musculus mRNA for high mobility group 2 protein (MOUSE);, mRNA sequence.
AA231198
NID 91853555
VERSION AA231198.1 GI:1853555
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE
AUTHORS Marra,M., Hallier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)

CONTACT: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800

Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:412711
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 23.

FEATURES

Source

1. 413

Location/Qualifiers

/organism="Mus musculus"
 /strain="C57BL/6j"
 /db_xref="taxon:10090"
 /map="788H12:14924.3"
 /clone="IMAGE:673007"
 /clone_lib="Soares mouse 3NME12 5"
 /sex="unknown"
 /tissue_type="fetus"
 /dev_stage="12.5dpc total fetus"
 /lab_host="DH10B"
 /note="Organ: whole fetus; Vector: pRT3D-Pac (Pharmacia)
 with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 1st strand cDNA was primed with a Not I - oligo(dT) primer
 15' TGTACCAATCTGAAGTGGAGCGCGCCCTATTATTTTATTTT
 3'], on total mouse RNA [provided by Minoru Ko, Wayne
 State Univ.]; double-stranded cDNA was ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of the modified pRT3 vector.
 Library went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

108 a 98 c 118 g 89 t

ORIGIN

Query Match 74.0%; Score 14.8; DB 30; Length 413;
 Best Local Similarity 65.0%; Pred. No. 2.1e+02;
 Matches 13; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARCA 20

DB 92 GGTGAAGACTTGCCCAAGA 111

*0

Search completed: September 13, 1999, 15:45:57
 Job time: 4045 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

Run on: September 13, 1999, 15:04:48 ; Search time 329.34 Seconds

(without alignments)
560,084 Million cell updates/sec

Title: US-09-325-095-1214.SEO

Sequence: 1 AARATGAGATGATGATTTTCA.....TTTCATATGATGATTTGGCCA 58

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database: GenEmbl.*

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sy:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_hgt:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_hgt1:*
35: gb_hgt2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	33	56.9	38	5	A37242	Sequence 11

2	31.8	54.8	3452	4	AF001076	AF001076 Gallus ga
3	31.8	54.8	3389	17	AF001075	AF001075 Avian sar
4	30.2	52.1	3207	3	BOVPHOS3KN	M93252 Bovine phop
5	30.2	52.1	3498	5	A37232	A37232 Sequence 1
6	30	51.7	1696	5	A17975	A17975 Glucose-6-p
7	28.6	49.3	3424	10	HSPH3K	Z29090 H.sapiens m
8	28.6	49.3	3207	10	HSU79143	U79143 Human phosp
9	26.8	46.2	1957	12	MMU03279	U03279 Mus musculu
10	26.8	46.2	3868	10	HSU110DEL	M64446 L.mesentero
11	26.6	45.9	3525	10	HSU57843	Y10055 H.sapiens m
12	26.6	45.9	3520	11	HSU68453	U57843 Human phosp
13	26.6	45.9	5220	7	SPC1672	U68453 Human phosp
14	26.4	45.5	36000	36	CEL721E12	AL031324 S.pombe c
15	25.8	44.5	35820	37	CEL2DNEINH	U06440 Caenorhabdi
16	25.8	44.5	35103	37	CEL2DNEINH	L33260 Caenorhabdi
17	25.6	44.1	2028	5	E03267	E03267 cDNA sequen
18	25.6	44.1	1347	5	E03268	E03268 cDNA sequen
19	25.6	44.1	2028	5	E05210	E05210 DNA encodin
20	25.6	44.1	1347	5	E05211	E05211 DNA encodin
21	25.6	44.1	2028	5	I72356	I72356 Sequence 1
22	25.6	44.1	1367	5	I72357	I72357 Sequence 3
23	25.6	44.1	2028	9	HUHPH8R1	D90187 Human mRNA
24	25.6	44.1	1367	9	HUHPH8R2	D90188 Human mRNA
25	25.6	44.1	1167	11	AF037351	AF037351 Homo sapi
26	25.4	43.8	95930	9	HS332011	Z98043 Human DNA s
27	25.4	43.8	22720	36	CEC34B7	Z83220 Caenorhabdi
28	25.2	43.4	197082	11	AC006249	AC006249 Homo sapi
29	25.2	43.4	23960	36	CEC54C8	Z83102 Caenorhabdi
30	25	43.1	128330	9	HS110P6	AL031056 Human DNA
31	25	43.1	81984	10	HSJ10E17	Z70272 Human DNA s
32	25	43.1	3132	12	MMU06587	U06587 Mus musculu
33	24.8	42.8	3213	10	S67334	S67334 phosphatidy
34	24.8	42.8	124351	11	AC005185	AC005185 Homo sapi
35	24.8	42.8	178707	35	AC007435	AC007435 Homo sapi
36	24.8	42.8	24159	42	AC005677	AC005677 Homo sapi
37	24.6	42.4	68726	7	AB017063	AB017063 Arabidops
38	24.6	42.4	84710	7	AB026643	AB026643 Arabidops
39	24.6	42.4	187755	34	AC006718	AC006718 Caenorhab
40	24.6	42.4	31310	36	CELPHE5	U08087 Caenorhabdi
41	24.6	42.4	3712	36	DMPHOS3K1	Y09070 D.melanogas
42	24.4	42.1	16450	1	MTYV051	AL022073 Mycobacte
43	24.4	42.1	2037	5	I41352	I41352 Sequence 1
44	24.4	42.1	119001	8	ATRC005896	AC005896 Arabidops
45	24.4	42.1	179599	11	AC006305	AC006305 Homo sapi

ALIGNMENTS

RESULT	1	A37242	38 bp	DNA	PAT	05-MAR-1997
LOCUS	A37242	Sequence 11 from Patent WO9403609.				
DEFINITION	A37242					
ACCESSION	g2294353					
VERSION	A37242.1	GI:2294353				
KEYWORDS						
SOURCE	unidentified.					
ORGANISM	unclassified.					
REFERENCE	1 (bases 1 to 38)					
AUTHORS	Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.					
TITLE	EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE					
JOURNAL	OR PROTEIN KINASE AND ASSAYS USING THEM					
COMMENT	Patent: WO 9403609-A 11 17-FEB-1994;					
FEATURES	IMP CANCER RES TRCH (GB)					
source	Other Publication JP 8503124T 960409.					
BASE COUNT	7 a	6 c	8 g	11 t	6 others	
ORIGIN	1..38					
	/organism="unidentified"					
	/db_xref="taxon:32644"					

Query Match 56.9%; Score 33; DB 5; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.11;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GATGATGGCCACACTGTTTCATATGATTTGGCCA 58
 1 GATGATGGCCACACTGTTTCATATGATTTGGCCA 35

RESULT 2
 AF001076 3452 bp mRNA VRL 08-JUL-1997
 LOCUS Gallus phosphoinositide 3-kinase catalytic subunit mRNA,
 DEFINITION complete cds.
 ACCESSION AF001076
 NID 92245505
 VERSION AF001076.1 GI:2245505
 KEYWORDS
 SOURCE Chicken.
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
 Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
 REFERENCE 1 (bases 1 to 3452)
 AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
 TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase
 JOURNAL Science 276 (5320), 1848-1850 (1997)
 MEDLINE 97334438
 REFERENCE 2 (bases 1 to 3452)
 AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
 TITLE Direct Submission
 JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
 FEATURES
 source 1..3452
 /organism="Gallus gallus"
 /db_xref="taxon:9031"
 /issue_type="brain"
 1..150
 /gene="c-p3k"
 /note="p3k proto-oncogene"
 151..3357
 /note="c-p3k"
 /codon_start=1
 /product="phosphoinositide 3-kinase catalytic subunit"
 /protein_id="AA062534.1"
 /db_xref="PID:92245506"
 /db_xref="GI:2245505"
 /translation="MPPRPSGELMGHILMPRIIVECLLPNGMIVTECLREATLLT
 IKHELFKARKYPLVQLQDESSYIFSVTOAEEREFEDTRRLCDLTFQPFKVI
 EGVGRREKILNREIGFALGMPCIEFDMVKDEPDRRIINLVCKEADLDANAP
 SNALVCPNVSSEPELKRITNKDKGIIIVYIVYSPNDKQYITIKIHDCVPE
 OVIAERIKRTSMLSLSEQLKCYLEVGKYLAVCGDELLKLYSYQYRASC
 IMLGRPNIMLAKESLYQLDPTFTMPSYSRISTAPYNGEATNSLMTINSAL
 RLICATYVNNIDIDIKIYVGTGIGHGEPICDVAHQPCSPNEMKLSYDM
 IPDLPRARLCLISICVGRKAKHEKPLANGINMEFYDTLVSGKALNLMVAPN
 GLEDLNPICVTGSPNPKETPCLELDFDMFSNPKPDMVTVEHANMTISRELQPNY
 SVAGSNRIARBNELRESKEDRLAICTDPISEITOEKDLKSHRHICVTPPELLP
 KLLSVKNMSRDEVAOMTCLVDMPPIKEQAMELLDCMYPDPAFAVRCLLEYLT
 DKLQYLIQVQLVKEYEQLDNLQVLRFLKRLANQRIQHFEFVHLKEMNKTVSQ
 RGLLESEYCRAGMYLKLISROVEMERKLINTDLKOEKDETOKVMKFLVQMR
 RFDMDALOGFISPLNPAHQNLRLKEGRIMSARPLMNMENDISLVEFQMR
 IIFKNGDDLRQDMLTQIIRIMENIMONGLDLRLMYPGCLSGDVCGLIEVRSHT
 IMQIOCKGKALQPNSTHTLQWLKDNKNGEMDAIDLFTRSCHGICVATFTLIGIG
 DRHNSINIMKDGQLFHDGFHDLKRRKFGYKRRVPEVLTQDFLIVISGADECT
 KTRERPEQEMCYKAVLAIRQANLFINFSMLSGMPELOSFDIAYIRKTLALDK
 TQDEALEYFMKMDNDAHGGMTTKMDMIRHTIKQNALN"
 3358..3452

BASE COUNT 1073 a 672 c 783 g 924 t
 ORIGIN

Query Match 54.8%; Score 31.8; DB 4; Length 3452;
 Best Local Similarity 81.1%; Pred. No. 0.36;
 Matches 30; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 22 ACATGATGGCCACACTGTTTCATATGATTTGGCCA 58
 1 ACATGATGGCCACACTGTTTCATATGATTTGGCCA 2957

RESULT 3
 AF001075 3389 bp mRNA VRL 08-JUL-1997
 LOCUS Avian sarcoma virus 16 gag-v-phosphoinositide 3-kinase catalytic
 DEFINITION subunit fusion protein (gag-v-p3k) and retroviral env mRNAs,
 partial cds.
 ACCESSION AF001075
 NID 92245502
 VERSION AF001075.1 GI:2245502
 KEYWORDS
 SOURCE Avian sarcoma virus 16.
 ORGANISM Avian sarcoma virus 16.
 REFERENCE 1 (bases 1 to 3389)
 AUTHORS Chang, H.W., Aoki, M., Furman, D., Auger, K.R., Bellacosa, A.,
 Tschilis, P.N., Cantley, L.C., Roberts, T.M. and Vogt, P.K.
 TITLE Transformation of chicken cells by the gene encoding the catalytic
 subunit of PI 3-kinase
 JOURNAL Science 276 (5320), 1848-1850 (1997)
 MEDLINE 97334438
 REFERENCE 2 (bases 1 to 3389)
 AUTHORS Chang, H.W., Aoki, M. and Vogt, P.K.
 TITLE Direct Submission
 JOURNAL Submitted (23-APR-1997) Molecular & Experimental Medicine, The
 Scripps Research Institute, 10550 N. Torrey Pines Road, Mail-drop
 BCC-239, La Jolla, CA 92037, USA
 FEATURES
 source 1..3389
 /organism="Avian sarcoma virus 16"
 /db_xref="taxon:60629"
 <1..3252
 /gene="gag-v-p3k"
 /note="fusion gene of gag and v-p3k"
 1..78
 /gene="gag-v-p3k"
 /note="derived from gag gene"
 <1..3252
 /gene="gag-v-p3k"
 /codon_start=1
 /product="gag-v-phosphoinositide 3-kinase catalytic
 subunit fusion protein"
 /protein_id="AA062532.1"
 /db_xref="PID:92245503"
 /db_xref="GI:2245503"
 /translation="QQLIRTAPESTLTTPGEIIRYLVDRKASHLMPRIIVECLLPNG
 MIVTECLREATLLT IKHELFKARKYPLVQLQDESSYIFSVTOAEEREFEDTRRLCDLTFQPFKVI
 EGVGRREKILNREIGFALGMPCIEFDMVKDEPDRRIINLVCKEADLDANAP
 SNALVCPNVSSEPELKRITNKDKGIIIVYIVYSPNDKQYITIKIHDCVPE
 OVIAERIKRTSMLSLSEQLKCYLEVGKYLAVCGDELLKLYSYQYRASC
 IMLGRPNIMLAKESLYQLDPTFTMPSYSRISTAPYNGEATNSLMTINSAL
 RLICATYVNNIDIDIKIYVGTGIGHGEPICDVAHQPCSPNEMKLSYDM
 IPDLPRARLCLISICVGRKAKHEKPLANGINMEFYDTLVSGKALNLMVAPN
 GLEDLNPICVTGSPNPKETPCLELDFDMFSNPKPDMVTVEHANMTISRELQPNY
 SVAGSNRIARBNELRESKEDRLAICTDPISEITOEKDLKSHRHICVTPPELLP
 KLLSVKNMSRDEVAOMTCLVDMPPIKEQAMELLDCMYPDPAFAVRCLLEYLT
 DKLQYLIQVQLVKEYEQLDNLQVLRFLKRLANQRIQHFEFVHLKEMNKTVSQ
 RGLLESEYCRAGMYLKLISROVEMERKLINTDLKOEKDETOKVMKFLVQMR
 RFDMDALOGFISPLNPAHQNLRLKEGRIMSARPLMNMENDISLVEFQMR
 IIFKNGDDLRQDMLTQIIRIMENIMONGLDLRLMYPGCLSGDVCGLIEVRSHT
 IMQIOCKGKALQPNSTHTLQWLKDNKNGEMDAIDLFTRSCHGICVATFTLIGIG
 DRHNSINIMKDGQLFHDGFHDLKRRKFGYKRRVPEVLTQDFLIVISGADECT
 KTRERPEQEMCYKAVLAIRQANLFINFSMLSGMPELOSFDIAYIRKTLALDK
 TQDEALEYFMKMDNDAHGGMTTKMDMIRHTIKQNALN"
 3358..3452

SCAGYCATFTLIGDRHNSNIMVKGDLGFHIDFGHFLDHKKKFGYKRRPVFLT
ODLIVISKAOECKTREPREFOMCHKAYLAIRQHANLFTNLSMGLSGMPELOS
FDDIAYIRKTLALDKTTEODALEYFKOMANDHGGWTTKMDITFTTIKOHANL"

misc-feature
/gene="gag-v-p3k"
/note="sequence of unknown origin"
85..3252
/gene="gag-v-p3k"
/note="derived from c-p3k gene"
3250..3370
3371..>3389
/codon_start=1
/product="retroviral env"
/protein_id="AA06253.1"
/db_xref="GI:2245504"
/db_xref="GI:2245504"
/translation="MARNS"

3'UTR
CDS

BASE COUNT 1071 a 655 c 751 g 912 t

ORIGIN

Query Match 54.8%; Score 31.8; DB 17; Length 3389;
Best Local Similarity 81.1%; Pred. No. 0.36;
Matches 30; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 22 AACGATGATGCCACRCCTTCATATGATTTGGCCA 58
1 ||||||| ||:|||||:||||:||||:||||| ||

Db 2816 AAGATGATGACAACTGTTCAATATTGACTTTGGCCA 2852

RESULT 4
BOVPHOS3K 3207 bp mRNA MAM 19-AUG-1992
LOCUS Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete
DEFINITION cds.
ACCESSION M93252
M93252.1 GI:163519
KEYWORDS phosphatidylinositol 3-kinase.
SOURCE Bos taurus cDNA to mRNA.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae;
Bovinae; Bos.
1 (bases 1 to 3207)
Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R.,
Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F.,
Hsuan, J.J., Courtnedge, S.A., Parker, P.J. and Waterfield, M.D.
Phosphatidylinositol 3-kinase: Structure and expression of the 110
kd catalytic subunit
Cell 70, 419-429 (1992)
92354059

JOURNAL
MEDLINE
FEATURES
source
1..3207
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AA03068.1"
/db_xref="PID:9163520"
/db_xref="GI:163520"

CDS
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AA03068.1"
/db_xref="PID:9163520"
/db_xref="GI:163520"

LOCATION/Qualifiers
1..3207
/organism="Bos taurus"
/db_xref="taxon:9913"
/cell_line="SGBAR-1"
1..3207
/EC_number="2.7.1.67"
/product="phosphatidylinositol 3-kinase"
/protein_id="AA03068.1"
/db_xref="PID:9163520"
/db_xref="GI:163520"

SHAGLSNRLARDNELRENDKEDOLARICTRDLSETTEODKDFIMSHRHVCTPIELP
KLILSVKNSNDEVAQOMCYLRDMPRIPEBAMELLDENVDPVARGAVACLELYT
DKLSQYLIQVQVAKTBOYLDNLVLRFLKALNIGFHFHLSSEHKNKTVSO
RFGLLLESYCRACGMVTLKLNROVAMKRLINLIDLOEKEDETQYOMFLVEOMR
RPDEMDALOGFLSPINPAHOLGNLEBECRLMSAKRPLAMNEMPDJMSLELPONE
IIFKNGDDLRODMTLQIIRIEMENWOGDLRLPYGCLISGCVLIEVRSHT
IMQIOCKGKLGALQFNSHTIHOWKDKNKEIYDAIDLFTRSCAGYCATFTLIGG
DRHNSNIMVKGDLGFHIDFGHFLDHKKKFGYKRRPVFLTODFLIVISKAOECKT
KTREREROMCYKAYLAIRQHANLFTNLSMGLSGMPELOSFDDIAYIRKTLALDK
TEODALEYFKOMANDHGGWTTKMDITFTTIKOHANL"

BASE COUNT 1028 a 581 c 680 g 918 t

ORIGIN

Query Match 52.1%; Score 30.2; DB 3; Length 3207;
Best Local Similarity 78.4%; Pred. No. 1.3;
Matches 29; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 22 AACGATGATGCCACRCCTTCATATGATTTGGCCA 58
1 ||||||| ||:|||||:||||:||||:||||| ||

Db 2771 AAGATGATGACAACTGTTCAATATTGATTTGGACA 2807

RESULT 5
A37232 3498 bp DNA PAT 05-MAR-1997
LOCUS A37232
DEFINITION Sequence 1 from Patent WO9403609.
ACCESSION A37232
M9294345
VERSION 92294345
KEYWORDS A37232.1 GI:2294345
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 3498)
Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
EUCARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
Patent: WO 9403609-A 1 17-FEB-1994;
IMP CANCER RES TECH (GB)
Other publications JP 8503124T 960409.
COMMENT Location/Qualifiers
1..3498
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="GI:2294346"
/db_xref="PID:92294346"
/db_xref="GI:2294346"

JOURNAL
MEDLINE
FEATURES
source
1..3207
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="GI:2294346"
/db_xref="PID:92294346"
/db_xref="GI:2294346"

CDS
1..3207
/EC_number="unidentified"
/product="unnamed protein product"
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="GI:2294346"
/db_xref="PID:92294346"
/db_xref="GI:2294346"

LOCATION/Qualifiers
1..3207
/organism="unidentified"
/db_xref="taxon:32644"
1..3207
/note="unnamed protein product"
/codon_start=1
/protein_id="CAA02284.1"
/db_xref="PID:e304808"
/db_xref="GI:2294346"
/db_xref="PID:92294346"
/db_xref="GI:2294346"

RESULT	6
A17975	
LOCUS	A17975 1696 bp DNA PAT 05-APR-1994
DEFINITION	Glucose-6-phosphate dehydrogenase gene seq. ID No.:1.
ACCESSION	A17975
NID	G512533
VERSION	A17975.1 GI:512533
KEYWORDS	.
SOURCE	Leuconostoc mesenteroides.
ORGANISM	Leuconostoc mesenteroides.
	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
	Leuconostoc.
REFERENCE	1 (bases 1 to 1696)
AUTHORS	Jarsch,M. and Lang,G.
TITLE	Cloning and overexpression of glucose-6-phosphate dehydrogenase of
	Leuconostoc dextranucis
JOURNAL	Patent:EP 0469523-A 1 05-FEB-1992;

```

CDS
      123..1583
            /organism="Leuconostoc mesenteroides"
            /db_xref="taxon:1245"
            /codon_start=1
            /transl_table=11
            /product="glucose-6-phosphate dehydrogenase"
            /protein_id="CA01360.1"
            /db_xref="PID:g512534"
            /db_xref="GI:512534"
            /translation="MVSEIKTLVTFEGGTGLAKRLKLYSPVNLNKKGYLQEHFAIVG
            TARQSLSDDEFKQLQVDSIKETDQQAQAEAFIAHFSRADVIDAASYGILKAIEE
            AATPEFDIGNRIIFYMSVAAPFEGGTIAKLKEGGLAEGYGRIMTERFGSYTAEE
            LOSLENAFDNDOLFRIIDHYLGEKVONIALRGPNIFDAAKNNDYIKNVQVTLAEY
            LGVERAGYVDTTGALIDIMNHTMOIYGWITAMKPESEFNKDIQAANAPAFNLKITY
            NEEENVNYFYVAGAGDTADMYKPYLPEADYPAKSKNNTFAGELQPLPMWEEVPTF
            VRSSRLAAKQTRVDIYFKAKSTFNFGSQDQAEYSLLIIDPKGIETKLAKSEVD
            FNNTIINDMVAESDEDKKNTPEPTERMTHDNNMGSGNFADWNGVSIAMKFDVATIAA
            YDAKAPALPEYTKSSMGEASDGLAEKMGDAMVFKG"
523 a      311 c      361 g      501 t

```

		Query Match	51.7%	Score 30;	DB 5;	Length 1696;
		Best Local Similarity	62.0%;	Pred. No. 1.4;	Mismatches	0;
		Matches	31;	Conservative	9;	Gaps 0;
Oy	1	AARAGAGATGGATHTTTCACACGATGATGCCCACTGTTCATATWGAAY	50			
Db	607	AAATGCATTGGAAAATGCATTGGATGATGACCACACTGTTCCGATTTGAC	656			
RESULT	7					
HSPHIK	HSPHIK	3424 bp	mRNA	PRI	24-AUG-1995	
LOCUS	H.sapiens mRNA for phosphatidylinositol 3-kinase.					
DEFINITION	H.sapiens mRNA for phosphatidylinositol 3-kinase.					
ACCESSION	Z28090					
NID	5472990					
VERSION	Z28090.1 GI:472990					
KEYWORDS	phosphatidylinositol 3-kinase.					
SOURCE	human.					
ORGANISM	Homo sapiens					
	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;					

CD5	CDRES	source	Location/Qualifiers
			1..3424
			/organism="Homo Sapiens"
			/db_xref="taxon:9606"
			/chromosome="3q25.3"
			/cell_line="K562"
			13..3219

```

/product="phosphatidylinositol 3-kinase"
/protein_id="CA82333.1"
/db_xref="P0:9472991"
/db_xref="GI:472991"
/db_xref="SWISS-Prot:P42336"
/translation="MPRRSSGELGWGHTMLPPRIIVECLLPNGMITYLECLREALTIVT
IKHLEFKARKYPLHQLLODESSYIFVSYQAEEREEFEPETRRLLDRLFOPLKAT
EAPGNREKILINNEGIAPGVCEPFGVDPVOFRRRIPLVNCSEAVDRLDPSNH
SRMAYVPPHVESPELPHKIIYKNDKGROIVIVIVYVSNNDKOKKTLINDQVSE
QVLAFAIRKTRBSMLTSSBOLKCYALEQSKYKTLISGDCGFLEYKPLSYQVYRSC
IMGRBNMLKMAKESITSDLPDCTFPMPEYSRITATPYNAGETSRKLYINRL
RILLCATVYNNLRIDDKLIVYRGTHGSEPLCDNVNTOVRVCSNRNMEVNTYDI
IPLPRLARALCISGVKGRGAKEEBCKPLAMWINLFDYDITLVGKALNMLPAPH
GLIDPLRLICVTSNPKETPCLEBEDFMSVYKPPDMVYIEEHNAVSVRSRAGSY
SHGLSRIARLNDRELRENDKEOLKAITRDPLEITIEOEKDFMRSHNYCVTLEILP
KLLSVKNSRDEVAQWQYCLYKMDPKPKPOAMELLDQNPDMVAGFAVRCLEKYL
DDLSITLYLNOVLYKEQYLDNLIVARELLAKLALNORIGHFFFNHLSKMAHKYTSQ
RFGILSTSRACAGSGLKHLNROYEAEMERKILNTDILQORDETSQVOKRFLVEQBR
RPFMDLQGLISPLNPAHQNGNLRLKECKINSARPLMLNNEPDIMSELLFQNNIE
IIRKNGDRLQMDLTLQILIRIMENIWONGDLRLMLPYGLSIGDCVGLLEVANSHT
IMQIOCGKGLKALQENSHTLHOMLKDKNKEIYDAIDFLDKSCAGCYVATEPIIGIG
DRHNSNINMYKDDGOLFIDHGFHLDHKKKKFKFGREVPYVLODFEIVYSKAOEET
KTEDEERFQEMCKAYVALIRQHANLIFNLFESMGLSGMPLESDFOIAYIRKTLALDK
TEDEALFYFKQNNDAHNGSWTKMKMDIIFTTIQHALLN"

```

Query Match	49.3%	Score 28.67	DB 10	Length 3424
Best Local Similarity	75.7%	Pred. No. 4.5		
Matches	28	Conservative	5	Mismatches 4; Indels 0; Gaps 0;
Q7	22	ACGATGATGCCACRCTCTTTCATATWATYTTGGCCA	58	
Db	2783	AAGACGATGCAACAACGTTCATATAGATYTTGGACA	2819	

RESULT	8
LOCUS	HSU79143
DEFINITION	Human phosphoinositide 3'-hydroxykinase p110-alpha subunit mRNA,
ACCESSION	U79143
NID	U79143.1
VERSION	GI:1763625
KEYWORDS	
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 3207)

[illegible]

FEATURES	source
JOURNAL	Mol. Cell. Biol. 14, 2675-2685 (1994)
MEDLINE	94187738
REFERENCE	2 (bases 1 to 3207)
AUTHORS	Klippel-Giese, A.
TITLE	Direct Submission
JOURNAL	Submitted (05-NOV-1993) Anke Klippel-Giese, CVRI and Dalich Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
source	Location/Qualifiers
	1..3207
	/organism="Mus musculus"
	/strain="Balb/c"
	/db_xref="taxon:10090"
	1..3207
	/standard_name="p110 protein"
	/codon_start=1
	/product="phosphatidylinositol 3-kinase 110 kDa subunit"
	/protein_id="AA18334.1"
	/db_xref="PID:g414995"
	/db_xref="GI:414995"
	/translation="MPRRSSGEIMGLHLPRIIVCELPNGMIVTLECRATVATIKHLEFARKYPLHOLDQETSYIFSVQAEAREEFPEETRIQLRFOELKVIYEPVGNREPKTLINREPTVIGTCMPCEPMDPVPDORRNILNVCRAVDRLNPSH SRAMVYPPVWSSPELPKRIYTKLDRGQIIIVYIYVTSNNDRKQTKLLINDCYBEQVAEATIRKTRKSLSEQLKLCVLEIYQKRIKVCGCDEFLEKPIQYVIRSCIMLRMPNMLMAKESLYSQLPIDSFMPYSRSRIATPYMGETSKSLMYINSLRIKILCATYVNVNIRIDIKIYVRGTIHGSEPLDQNVNTOVRCSNRMENMYDIYIPDLPRARLCISGICSGKRGAKKEHCPLAMGNINLFDYDTLIVSGKMLNLPVPHGLDILNPLIGVTSNPKKPKCELEPEDFWSSVYKPRDMSVIEHNAKVSREGEYSHTGSLNRRLRDLRENDKEQLRALDTRDPLSEITDEKDFLMSHRHYCVTIPELPDKLISVKNSSRDEVAQMYCLVMDPPIKPRQAAELLDCA NPDVWRSFAVRCLEKYLTDLKLSQYLQIVLVAKYEQYLDMLNVLFLKRLATNORIGHFEFWHLKSMHNKTVSQRFGLISLSCYRACAGMYKHLNRQVEAKELINTDLKQKKNEDTQVOKKFLVEQWRQPDMDLQGLFSLPNAHQLGNRLRECGIMSARKPLKMDENPTQINSELLFQNNELIRKNGDRLQMDTLQITIRIMENIMQGLDRLMPLYGSLISDCGQILEVRSNFTIMQIQCKGALQENSHRLHOMLKKRNGEITDAIDITFTSCAGTCVATITLIGIGDRHNSINMAGVDQLEFHDHGFPLDHRKKRGYKREVPVLVQDFLIVSKAQEYTKTRFERFOEMCYKAVYLAIRQHANLFTNLFSMIMGSMPLQSEFDIYAIRKTLADKTEQEALEYFTKOMNDAHNGWTKMDIIFTIKQHALN"
BASE COUNT	958 a 693 c 734 g 822 t
ORIGIN	
Query Match	49.3%; Score 28.6; DB 12; Length 3207;
Best Local Similarity	75.7%; Pred. No. 4.5;
Matches	28; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
0Y	22 ACGATGATGCCARCTGTTCAATGATGATTTGGCCA 58 1 11111 11 111111111111111111 11
Db	2771 AAGATGACGACAGCTGTTTCATATAGATTTTGGCA 2807
RESULT 10	
LEUGEDP	LEUGEDP 1957 bp DNA circular BCT 31-MAR-1994
LOCUS	L.mesenteroides glucose-6-phosphate dehydrogenase gene, complete cds.
DEFINITION	
ACCESSION	M64446
NID	g149630
VERSION	M64446.1 GI:149630
WORDS	glucose-6-phosphate dehydrogenase.
SOURCE	L.mesenteroides DNA.
ORGANISM	Leuconostoc mesenteroides Bacteria: Firmicutes; Bacillus/Clostridium group; Lactobacillaceae; Leuconostoc. 1 (bases 1 to 1957) Lee, W.T., Flynn, T.G., Lyons, C. and Levy, H.R. Cloning of the gene and amino acid sequence for glucose 6-phosphate dehydrogenase from leuconostoc mesenteroides J. Biol. Chem. 266, 13028-13034 (1991) 91302321 Location/Qualifiers 1..1957

supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPBC25H2.01c, SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand). The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid c1672 is overlapped by cosmid c1183 at its 3' and by cosmid c962 at its 5'.

FEATURES

```

source
    location="Qualifiers"
    1..36000
    /organism="Schizosaccharomyces pombe"
    /strain="972h"
    /db_xref="taxon:4896"
    /chromosome="III"
    /clone="cosmid c1672"
    /map="III"
    /complement(1..1848)
    /note="nominal overlap with c962 S. pombe chromosome 3"
    1579..2789
    /gene="SPCC1672.01"
    join(1579..1610,1727..1853,2028..2789)
    /gene="SPCC1672.01"
    /note="SPCC1672.01, len:306aa, similarity: yeast, YFR025C, HIS9_YEAST, histidinol-phosphatase, (335aa), fasta scores, opt:306, E():0, (33.3% identity in 330 aa overlap)"
    /codon_start=1
    /product="histidinol-phosphatase"
    /protein_id="CA20439.1"
    /db_xref="pid:e1316727"
    /db_xref="gi:3451302"
    /translation="MPISSHSGGRCILAGKRLDVTQIAIOGROSRFPEHPRD
RVEDLPRELHQPEDLFRTPEDYVNEARLKNQNGDQINILGATETIRRESVLL
KSLNTRYNLIDYVGVSHVNSIPIDSPFLQKALQHVNNBQELIDVFIEQYDLMQ
RHAPLVIGHFDLICEFAPDAKVEFNKSNVWELIQRNIRKAVSYGGEIETNSFRK
GKRTAVPQORLELWVEQAGQLSDSDSGPQOVGLNYHLAKSYLDKCGTSLCMLEK
GPDNGTVYKNKTVDMVMSKFGTNGITNT"
    1611..1616
    /misc_feature
    /gene="SPCC1672.01"
    /note="gtaagt, splice donor sequence"
    1711..1726
    /misc_feature
    /gene="SPCC1672.01"
    /note="ctaatctttatgaag, splice branch and acceptor"
    1854..1859
    /misc_feature
    /gene="SPCC1672.01"
    /note="gtaagt, splice donor sequence"
    2009..2027
    /misc_feature
    /gene="SPCC1672.01"
    /note="ctaatctatcatcttag, splice branch and acceptor"
    complement(3865..4629)
    /gene="sep1"
    /complement(3865..4629)
    /gene="sap1"
    /note="SPCC1672.02c, len:254aa"
    /codon_start=1
    /product="switch-activating protein 1"
    /protein_id="CA20440.1"
    /db_xref="pid:e1316728"
    /db_xref="pid:g3451303"
    /db_xref="gi:3451303"
    /translation="MEAPKRELKSYRKKNASLSPSSSPAKAQRTHLSLEKIKLRLY
VHHKHLNDRKTSSEPAKIRAGIEDEGAIATBESRCNOISIMVAYQVQRAHQPG
MKTTPEDDLQDCDEWKARLSELQYRKRFVGRKKCCNDSEINERKLTLEEQQNV
DKLVAKVFLSKHLHDNEERKLMQVNAKMEVLAENRRLQDLDDHDLLSKLEPPSAVA
PGVNGVTMGNGMGNMNAINGLHSSISIPNLGDH"
    complement(7403..8986)
    /gene="SPCC1672.03c"
    /misc_feature
    /gene="SPCC1672.03c"
    /note="SPCC1672.03c, len:527aa, similarity: YDL238C, YD3C_YEAST, hypothetical 55.2 kd protein, (489aa), fasta scores, opt:1131, E():0, (43.6% identity in 473 aa overlap)"
    /codon_start=1
    /product="hypothetical protein"
    /protein_id="CA20441.1"
    /db_xref="pid:e1316729"
    /db_xref="pid:g3451304"
    /db_xref="gi:3451304"
    /translation="NDGRCQVFCVKLIHPISIGLETITDAPVGYNKKIYFLKSM
PTLEAKSHHLKATITHKLPIDQFMFPGILDIHTARQVPSNGIGIDVPLQMLEK
YTFPLESLADLEEARQVYKRVETRLSNGTFASVETLHPPTSALAEICYSYGR
ATVIGCKNNMNSLDPEYCKERSAESLEAROIISYMSIDPKREMTPTITPFAFSC
EDLISGCELEAKNHLPIOTHSENTSIIELELFPKRSYADVYKLTLPOTIL
AAHILDEBEITELFKRSSGISHCPTSILASGLANRKILDSGINGLDVGSGY
APSLIILRHAMTFRSLSYVGDPRKVLDSILLIYLAQGAPEVSRGDSFRAG
KYWDALYDLSAEHRSQVDFERDTPWMLSKWFTSDRNLAQVWVAGRLVSGFEK
ANLKNSTPLTNGVTSVSGHQVPEKLTQAHLLPRTQCVDPPSCCGGCKEBSRCRENC
KGAYPANATVVEEDSGMS"
    complement(10212..10584)
    /gene="SPCC1672.04c"
    complement(join(10212..10465,10533..10584))
    /gene="SPCC1672.04c"
    /note="SPCC1672.04c, len:101aa"
    /codon_start=1
    /product="very hypothetical protein"
    /protein_id="CA20442.1"
    /db_xref="pid:e1316730"
    /db_xref="pid:g3451305"
    /db_xref="gi:3451305"
    /translation="MTTRPEPERSGFPDLHGECHTWNKQYLECIKVRNOEBCRL
AKKYLQCRMDTGLFGKDMKNIGFSGDENASTLSLSSNDGNMNSSSDNKTGGE"
    complement(10466..10483)
    /gene="SPCC1672.04c"
    /note="ctaatctgtctgag, splice branch and acceptor"
    complement(10527..10532)
    /gene="SPCC1672.04c"
    /note="gtaagt, splice donor sequence"
    complement(11755..12960)
    /gene="SPCC1672.05c"
    complement(11755..12960)
    /gene="SPCC1672.05c"
    /note="SPCC1672.05c, len:401aa"
    /codon_start=1
    /product="tyrosyl-tRNA synthetase cytoplasmic EC 6.1.1.1"
    /protein_id="CA20443.1"
    /db_xref="pid:e1316731"
    /db_xref="pid:g3451306"
    /db_xref="gi:3451306"
    /translation="MSLTPDERKYELITRNLOEVLAGKMRITELNDRDISLYGSAPTG
RPHCGYFVPMKLADFLKAGVHYVLADVAFLDMKAPMELVOHVRVETIYKSI
LKSINPIELKRETVSGSYOLKEDYTLDNFLASTVLRBAGAEVYKQVAPSLAS
SLIVRGMALDEQVYGVDFGQVDRITIPMAEVLVLYGKKRGLHMANVPYSLAG
GMSASAANKIDILDDPKTVNKRIRSAFCEPVEVNEGCLAFKIVSFPAEMLRGE
TEVYIDRPEQFEGGRVVFTEVEVAYADLISPDKLIGLESSVNTLLAIVQEDLQ
YPDIDSLIKAAVPPDKDLKAMKOKKODKRNKAKGTNASTIAESGAAPTESOTAE
SFKKILND"
    complement(11875..12930)
    /gene="SPCC1672.05c"
    /note="Pfam match to entry PF00579 tRNA-synt_1b, tRNA synthetases class I (Tyr and Tyr)"
    complement(13222..15984)
    /gene="SPCC1672.06c"
    complement(13222..15984)
    /gene="SPCC1672.06c"
    /note="SPCC1672.06c, unknown, len:920aa, similarity: S. cerevisiae, YLR410W, Q06685, chromosome xii cosmid 9931.7, (1146aa), fasta scores, opt:2648, E():0, (52.5% identity in 962 aa overlap)"
    /codon_start=1

```

CDS

gene

CDS

misc_feature

misc_feature

CDS

gene

misc_feature

gene

CDS

```

complement(7403..8986)
/gene="SPCC1672.03c"
/note="SPCC1672.03c, len:527aa, similarity: YDL238C, YD3C_YEAST, hypothetical 55.2 kd protein, (489aa), fasta scores, opt:1131, E():0, (43.6% identity in 473 aa overlap)"
/codon_start=1
/product="hypothetical protein"
/protein_id="CA20441.1"
/db_xref="pid:e1316729"
/db_xref="pid:g3451304"
/db_xref="gi:3451304"
/translation="NDGRCQVFCVKLIHPISIGLETITDAPVGYNKKIYFLKSM
PTLEAKSHHLKATITHKLPIDQFMFPGILDIHTARQVPSNGIGIDVPLQMLEK
YTFPLESLADLEEARQVYKRVETRLSNGTFASVETLHPPTSALAEICYSYGR
ATVIGCKNNMNSLDPEYCKERSAESLEAROIISYMSIDPKREMTPTITPFAFSC
EDLISGCELEAKNHLPIOTHSENTSIIELELFPKRSYADVYKLTLPOTIL
AAHILDEBEITELFKRSSGISHCPTSILASGLANRKILDSGINGLDVGSGY
APSLIILRHAMTFRSLSYVGDPRKVLDSILLIYLAQGAPEVSRGDSFRAG
KYWDALYDLSAEHRSQVDFERDTPWMLSKWFTSDRNLAQVWVAGRLVSGFEK
ANLKNSTPLTNGVTSVSGHQVPEKLTQAHLLPRTQCVDPPSCCGGCKEBSRCRENC
KGAYPANATVVEEDSGMS"
complement(10212..10584)
/gene="SPCC1672.04c"
complement(join(10212..10465,10533..10584))
/gene="SPCC1672.04c"
/note="SPCC1672.04c, len:101aa"
/codon_start=1
/product="very hypothetical protein"
/protein_id="CA20442.1"
/db_xref="pid:e1316730"
/db_xref="pid:g3451305"
/db_xref="gi:3451305"
/translation="MTTRPEPERSGFPDLHGECHTWNKQYLECIKVRNOEBCRL
AKKYLQCRMDTGLFGKDMKNIGFSGDENASTLSLSSNDGNMNSSSDNKTGGE"
complement(10466..10483)
/gene="SPCC1672.04c"
/note="ctaatctgtctgag, splice branch and acceptor"
complement(10527..10532)
/gene="SPCC1672.04c"
/note="gtaagt, splice donor sequence"
complement(11755..12960)
/gene="SPCC1672.05c"
complement(11755..12960)
/gene="SPCC1672.05c"
/note="SPCC1672.05c, len:401aa"
/codon_start=1
/product="tyrosyl-tRNA synthetase cytoplasmic EC 6.1.1.1"
/protein_id="CA20443.1"
/db_xref="pid:e1316731"
/db_xref="pid:g3451306"
/db_xref="gi:3451306"
/translation="MSLTPDERKYELITRNLOEVLAGKMRITELNDRDISLYGSAPTG
RPHCGYFVPMKLADFLKAGVHYVLADVAFLDMKAPMELVOHVRVETIYKSI
LKSINPIELKRETVSGSYOLKEDYTLDNFLASTVLRBAGAEVYKQVAPSLAS
SLIVRGMALDEQVYGVDFGQVDRITIPMAEVLVLYGKKRGLHMANVPYSLAG
GMSASAANKIDILDDPKTVNKRIRSAFCEPVEVNEGCLAFKIVSFPAEMLRGE
TEVYIDRPEQFEGGRVVFTEVEVAYADLISPDKLIGLESSVNTLLAIVQEDLQ
YPDIDSLIKAAVPPDKDLKAMKOKKODKRNKAKGTNASTIAESGAAPTESOTAE
SFKKILND"
complement(11875..12930)
/gene="SPCC1672.05c"
/note="Pfam match to entry PF00579 tRNA-synt_1b, tRNA synthetases class I (Tyr and Tyr)"
complement(13222..15984)
/gene="SPCC1672.06c"
complement(13222..15984)
/gene="SPCC1672.06c"
/note="SPCC1672.06c, unknown, len:920aa, similarity: S. cerevisiae, YLR410W, Q06685, chromosome xii cosmid 9931.7, (1146aa), fasta scores, opt:2648, E():0, (52.5% identity in 962 aa overlap)"
/codon_start=1

```


CC These sequences were transformed into Schizosaccharomyces pombe cells
 CC in an embodiment of the invention. In the presence of thiamine the
 CC promoter is inactive and the cells carrying the PKC plasmids grow as
 CC the parental strain. In the absence of thiamine the nmt promoter
 CC functions and the PKC is induced. PKC activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SQ Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 56.9%; Score 33; DB 1; Length 38;
 Best Local Similarity 100.0%; Pred. No. 0.00091;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 DB 1 GATGATGGCCARCTGTTTCATATGATGATTTGGCCA 35

RESULT 3
 ID 051155 standard; CDNA; 3207 BP.
 AC 051155;
 DT 12-APR-1994 (first entry)
 DE P110 CDNA.
 KW phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; propylaxis; therapy;
 KW platelets; neurophil activity; 3-phosphorylated phosphoinositides;
 KW ss.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /*tag- a
 FT /note= "PI3- kinase p110"
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parterp], Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 DR P-PSDB; R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English.
 CC Southern blot analysis was performed using a bovine CDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human CDNA isolated
 CC from a CDNA library constructed from mRNA isolated from the human
 CC cell line K61a. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (ant)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC propylaxis or therapy. Platelet or neurophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 SQ Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 52.1%; Score 30.2; DB 1; Length 3207;
 Best Local Similarity 78.4%; Pred. No. 0.043;
 Matches 29; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 22 AGCATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 DB 2771 AGATGATGGACACACTGTTTCATATGATGATTTGGACA 2807

RESULT 4
 ID 057012 standard; CDNA to mRNA; 3498 BP.
 AC 057012;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase 110 kD catalytic subunit CDNA.
 KW 110 kD catalytic subunit; phosphatidy] inositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /*tag- a
 FT /product= p110

PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-01654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 DR P-PSDB; R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure; Fig 1; 71pp; English.
 CC This sequence encodes the 110 kD catalytic subunit of the phosphatidy] inositol (Ptdins) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the Ptdins
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the Ptdins 3-kinase
 CC catalytic subunit is induced. Ptdins activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SQ Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 52.1%; Score 30.2; DB 1; Length 3498;
 Best Local Similarity 78.4%; Pred. No. 0.045;
 Matches 29; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

OY 22 AGCATGATGGCCARCTGTTTCATATGATGATTTGGCCA 58
 DB 2771 AGATGATGGACACACTGTTTCATATGATGATTTGGACA 2807

RESULT 5
 ID 020959 standard; DNA; 1696 BP.
 AC 020959;
 DT 12-MAY-1992 (first entry)
 DE L-dextranase glucose-6-phosphate dehydrogenase coding sequence.
 KW G6P-DH; glycolysis; enzyme; DSM 20187; thermal stability; ss.
 OS Leuconostoc dextranans.
 FH Key Location/Qualifiers
 FT cds 1..122
 FT promoter
 FT /*tag- a
 FT /*tag- b
 FT /product= G6P-DH
 PN EP-469523-A.
 PD 05-FEB-1992.
 PF 29-JUL-1991; 112739.
 PR 30-JUL-1990; DE-024158.
 PA (BOE) BOEHRINGER MANNHEIM GMBH.
 PI Jarsch M, Lang G;

RESULT
X20017/0

glucose-6-phosphate. The kinetic

CC sequence is a dimer with identical subunits of mol. wt. 54,800. It
CC can utilize either NAD^+ or NADP^+ as coenzyme in the oxidation of
CC glucose-6-phosphate. The kinetic mechanism differs depending on
CC whether NAD^+ or NADP^+ is used in the reaction. Binding of NAD^+

CC is designated p110-delta. The protein is expressed selectively in
 CC white blood cells and melanomas. The protein is a lipid kinase with
 CC broad phosphoinositide specificity and specific tissue localisation,
 CC possibly involved in regulation of melanoma metastases (it is not
 CC expressed in normal melanocytes). Fragments of the present
 CC sequence are used to detect tissue-specific expression, and a
 CC similar analysis can be done at the protein level using antibodies in
 CC standard immunoassays. These assays are particularly used for diagnosing
 CC and predicting motility/invasiveness of metastatic cancer cells. The
 CC protein can be used in human or veterinary medicine for controlling
 CC motility of cells, where the protein increases motility while antisense
 CC sequences are used to reduce it.
 CC Sequence 3387 BP: 711 A; 1044 C; 1001 G; 631 T;

Query Match 45.9%; Score 26.6; DB 1; Length 3387;
 Best Local Similarity 74.3%; Pred. No. 0.95;
 Matches 26; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 24 GATGATGGCCARCTGTTTCATATGATTTGGCCA 58
 DB 2707 GAGAGTGGCAGCTGTTCCACATTTGTTGGCCA 2741

RESULT 15
 ID V31340 standard; cDNA: 5220 BP.
 AC V31340;
 DT 12-OCT-1998 (first entry)
 DE Human phosphatidylinositol 3-kinase p110 delta subunit cDNA.
 KW Phosphatidylinositol 3-kinase; p110 delta; human; immune system;
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 196..3350
 FT /tag= a
 PN WO9823760-A1.
 PD 04-JUN-1998.
 PF 25-NOV-1997: U21655.
 PR 25-NOV-1996: US-777405.
 PA (ICOS-) ICOS CORP.
 PI Chantry DH, Hoeckstra MF, Holtzman DA;
 DR WPI: 98-322736/28.
 P-PSDB: W58570.
 PT New phosphatidylinositol 3-kinase catalytic subunit - used to
 PT develop products for modulating kinase activity in immune system
 PT signalling and in carcinogenesis
 PS Claim 4: Page 27-33; 53pp: English.
 CC This full-length composite cDNA encodes the p110 delta catalytic
 CC subunit (see W58570) of human phosphatidylinositol 3-kinase. It
 CC was assembled from clone #249 obtained by PCR amplification (see
 CC W31341-44) of human peripheral blood mononuclear cell cDNA, clone
 CC #928 obtained by screening a human macrophage cDNA library, and
 CC further clones obtained by RACE and PCR (see V31345-50). The
 CC following are claimed: (1) a purified and isolated polynucleotide
 CC (PN) encoding p110 delta; (2) a vector comprising a DNA as in (1);
 CC (3) a host cell stably transformed or transfected with a DNA as in
 CC (1); (4) PN encoding a lipid kinase, and hybridising to PN having
 CC the 5220 bp sequence; (5) a purified and isolated p110 delta
 CC polypeptide as in (4); (6) an antibody specifically immunoreactive
 CC with p110 delta; (7) a hybridoma cell line (especially 208F
 CC (HB 12200) producing a monoclonal antibody as in (6); and (8) a
 CC humanised antibody as in (6). p110 delta has kinase activity and
 CC may play a role in PI 3-kinase mediated signalling in the immune
 CC system and in carcinogenesis. The products can be used to develop
 CC agents that modulating p110 delta kinase activity and to develop
 CC diagnostic reagents (claimed). They may also be used for detection
 CC and diagnosis of p110 delta in a biological sample.
 SQ Sequence 5220 BP: 1120 A; 1525 C; 1483 G; 1092 T;

Query Match 45.9%; Score 26.6; DB 1; Length 5220;
 Best Local Similarity 74.3%; Pred. No. 1.1;

Matches 26; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

OY 24 GATGATGGCCARCTGTTTCATATGATTTGGCCA 58
 DB 2902 GAGAGTGGCAGCTGTTCCACATTTGTTGGCCA 2936

Search completed: September 13, 1999, 16:21:48
 Job time: 4008 sec

TITLE
Bednarik, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dime, D., Feng, D.-F., Fertile, A., Fischer, C., Hastings, G.A., He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K., Kozak, D.L., Kunsch, C., Hung, J., Li, H., Melsner, P.S., Olsen, H., Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C., Fraser, C.M., and Venter, J.C.
JOURNAL
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
MEDLINE
Nature 377 (6547 Suppl), 3-174 (1995)
COMMENT
On May 5, 1995 this sequence version replaced g1:798292.

FEATURES
source
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
Seq primer: M13 Reverse.
Location/Qualifiers
1..243
/organism="Homo sapiens"
/db_xref="ATCC (inhost):188107"
/db_xref="taxon:9606"
/map="19"
/clone_1lb="Thymus II"
/sex="male"
/dev_stage="adult, 19 yrs"
/note="Organ: thymus; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
EcoRI: Site_2: XhoI"

BASE COUNT
74 a 63 c 37 g 69 t

ORIGIN

Query Match
Best Local Similarity 40.3%; Score 23.4; DB 33; Length 243;
Matches 28; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
3 RATGAGTGGATHTTTCAYACAGATGCGCCTTCATTCATTCATTCG 55
72 AATGGTACGTATTCCTTTGAGGTGATGACGTCTCAATGATTCG 20

RESULT 15
AA951971 584 bp mRNA EST 24-NOV-1998
LOCUS LD28407.5prime LD Drosophila melanogaster embryo pot2 Drosophila
DEFINITION melanogaster cDNA clone LD28407 5prime similar to AF017777:
Drosophila melanogaster tweety (tty), flightless (fli), dodo (dod),
penguin (pen), small optic lobes (sol), innocent bystander (lby),
wacław (waw), bobby sox (bbs), slugish (slg), helicase (hlc),
misto (mst), and la costa (lcs) genes, mRNA sequence.
AA951971
ACCESSION 93114208
NID AA951971.1 GI:3114208
VERSION
KEYWORDS
SOURCE EST.
ORGANISM fruit fly,
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 584)
AUTHORS Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Brookstein, P., Lewis, S. and Rubin, G.M.
TITLE BDF/HMT Drosophila EST Project
JOURNAL Unpublished (1997)
COMMENT On Sep 29, 1997 this sequence version replaced g1:1520646.

CONTACT: Harvey, D.,
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: <http://www.fruitfly.org/EST/estefruitfly.berkeley.edu>
Plate: 284 row: A column: 7
High quality sequence stop: 520.

FEATURES
source
1..584
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="LD28407"
/clone_1lb="LD Drosophila melanogaster embryo pot2"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="XLI Blue"
/note="Organ: embryo; Vector: pot2; Site_1: EcoRI; Site_2: XhoI; Sized fractionated cDNAs were directly ligated into pot2."
BASE COUNT
165 a 152 c 180 g 86 t 1 others
ORIGIN

Query Match
Best Local Similarity 40.3%; Score 23.4; DB 40; Length 584;
Matches 28; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
5 TCGATGATHTTTCAYACAGATGCGCCTTCATTCATTCATTCG 57
DB 514 TCGATGCGACATCTCCACGCTGAGGAGACAGCTTCACAGATTCGCC 462

Search completed: September 13, 1999, 16:13:49
Job time: 5448 sec

JOURNAL confers than to flowering plants
REFERENCE 3 (bases 1 to 1547)
AUTHORS Theissen, G.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1999) Theissen G., Molecular Plant Genetics,
Max-Planck-Institut for Breeding Research, Carl-von-Linne Weg 10,
50829 Koeln, GERMANY

FEATURES
source
1. 1547
/organism="Gnetum gnemon"
/sub_species="gnemon"
/db_xref="taxon:3382"
159..933
/gene="gsm11"
159..933
/gene="gsm11"
/codon_start=1
/product="putative MADS domain transcription factor GGM11"
/protein_id="CA844457.1"
/db_xref="PID:e1487020"
/db_xref="PID:g5019460"
/db_xref="GI:5019460"
/translation="MGRGVRLKRIKINROYTFKRRNGLLKRAYELSYCLDAEVA
LIFFSRGKLYEFGSAGLTLEKRYOKCSVALQESNSDROATMHEVSKIKTYKEI
LORSORHLGEDIKPISTIRELOTLEKQIEVALTOVARKTQVAMDMADLKKEKRLQ
EVNKSRLKLDIEGVVSNALQAPPEPMSNALANVYALPTPPQNAVDEPTCK
LGIMLLSQAFPMILRTTYRGMSNLIIPANNQ"

BASE COUNT 465 a 308 c 372 g 402 t

ORIGIN

Query Match 62.1%; Score 36; DB 8; Length 1547;
Best Local Similarity 100.0%; Pred. No. 0.0085;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 23 TGACTGAGTCGACATCGATTGTTTTTTTTTTT 58
|||||
Db 1215 TGACTGAGTCGACATCGATTGTTTTTTTTTT 1180

RESULT 2
LOCUS XLU88561 638 bp mRNA VRT 16-MAY-1997
DEFINITION Xenopus laevis E2 ubiquitin conjugating enzyme (Ubc9) mRNA,
complete cds.
ACCESSION U88561
NID 92078330
VERSION U88561.1 GI:2078330
KEYWORDS African clawed frog.
SOURCE Xenopus laevis
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae;
Xenopus.
REFERENCE 1 (bases 1 to 638)
AUTHORS Saitoh, H., Pu, R., Cavenagh, M. and Dasco, M.
TITLE RanBP2 associates with Ubc9p and a modified form of RanGAP1
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 94 (8), 3736-3741 (1997)
MEDLINE 97268639
REFERENCE 2 (bases 1 to 638)
AUTHORS Saitoh, H., Pu, R., Cavenagh, M. and Dasco, M.
TITLE Direct Submission
JOURNAL Submitted (05-FEB-1997) LME/NICHD, NIH, Bldg. 18, Rm. 101, Bethesda,
MD 20892, USA
FEATURES
source
1. 638
/organism="Xenopus laevis"
/db_xref="taxon:8335"
/cell_type="oocyte"
/gene="Ubc9"
107..583
107..583

/gene="ubc9"
/note="Ubc9p"
/codon_start=1
/product="E2 ubiquitin conjugating enzyme"
/protein_id="AAB57736.1"
/db_xref="PID:92078331"
/db_xref="GI:2078331"
/translation="MSGIALSLQAEKAMKRDHPFGVVAFTKPNPQGTMLIMNECA
IPKRGTPWEGGELKRLMFLKDDYPSPPCKKFEPLFLHPNVPSPGIVCLSTLEEDKD
MRPATIKQILGIELNEPNIODPAQAEAYTICQNRVEYERKRAQAKKAPS"

Query Match 60.3%; Score 35; DB 4; Length 638;
Best Local Similarity 100.0%; Pred. No. 0.018;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 58
|||||
Db 638 GACTCGAGTCGACATCGATTGTTTTTTTTTT 604

RESULT 3
LOCUS A24382 35 bp DNA PAT 21-MAR-1995
DEFINITION dltI adaptor primer.
ACCESSION A24382
NID 9904477
VERSION A24382.1 GI:904477
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 35)
AUTHORS Samientos, P., De Taxis du Poet, P., Nitli, G. and Socheri, E.
TITLE Anti-thrombin polypeptides
JOURNAL Patent: EP 0501821 A 22 02-SEP-1992;
FARMITALIA CARLO ERBA S.r.L.
FEATURES
source
1. 35
/organism="synthetic construct"
/db_xref="taxon:32630"

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTGTTTTTTTTTTT 35

RESULT 4
LOCUS A27645 35 bp DNA PAT 04-JUN-1995
DEFINITION Hybrid dltI7-adaptor primer.
ACCESSION A27645
NID 91248481
VERSION A27645.1 GI:1248481
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 35)
AUTHORS POLYPHENOL OXIDASE GENES
TITLE Patent: WO 9302195-A 1 04-FEB-1993;
JOURNAL Location/Qualifiers
FEATURES
source
1. 35

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN /organism="synthetic construct"
/db_xref="taxon:32630"

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 5
LOCUS A37244 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 13 from Patent WO9403609.
ACCESSION A37244
NID 92294355
VERSION A37244.1 GI:2294355

KEYWORDS
SOURCE
ORGANISM

REFERENCE 1 (bases 1 to 35)

AUTHORS Parker, P. J., Goode, N. T., Nurse, P. M. and Waterfield, M. D.
TITLE EDUCAROTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
*OR PROTEIN KINASE AND ASSAYS USING THEM
JOURNAL Patent: WO 9403609-A 13 17-FEB-1994;
TMP CANCER RES TECH (GB)
Other publication JP 8503124T 960409.
COMMENT Location/Qualifiers

FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 6
LOCUS A40373 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 29 from Patent WO9425489.
ACCESSION A40373
NID 92296422
VERSION A40373.1 GI:2296422

KEYWORDS
SOURCE
ORGANISM

REFERENCE 1 (bases 1 to 35)

AUTHORS Monapatre, S. S. and Sehon, A. H.
TITLE DOWNREGULATION OF ALLERGEN-SPECIFIC IMMUNE RESPONSE
JOURNAL Patent: WO 9425489-A 29 10-NOV-1994;
UNIV MANITOBA (CA)

COMMENT Other publication AU 6674094 941121.
FEATURES
source 1..35
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 7
LOCUS A42335 35 bp DNA PAT 05-MAR-1997
DEFINITION Sequence 7 from Patent WO9502057.
ACCESSION A42335
NID 92297812
VERSION A42335.1 GI:2297812

KEYWORDS
SOURCE
ORGANISM

REFERENCE 1 (bases 1 to 35)

AUTHORS Gusterson, B. A., Crompton, M. R., Mitchell, P. J., Barker, K. T.,
Kamaleri, T., Page, M. J. and Spence, P.
TITLE PROTEIN TYROSINE KINASE AND LIGANDS THEREOF
JOURNAL Patent: WO 9502057-A 7 19-JAN-1995;
CANCER RES INST (GB)
Other publication AU 7080994 950206.
COMMENT Location/Qualifiers

FEATURES
source 1..35
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTCCTTTTTCCTTTT 35

RESULT 8
LOCUS A42384 35 bp DNA PAT 06-MAR-1997
DEFINITION Sequence 7 from Patent WO9502187.
ACCESSION A42384
NID 92297858
VERSION A42384.1 GI:2297858

KEYWORDS
SOURCE
ORGANISM

REFERENCE 1 (bases 1 to 35)

AUTHORS Gusterson, B. A., Crompton, M. R., Mitchell, P. J., Barker, K. T.,
Martindale, J. E., Page, M. J. and Spence, P.
TITLE CELL GROWTH FACTOR RECEPTORS
JOURNAL Patent: WO 9502187-A 7 19-JAN-1995;
CANCER RES INST (GB)
Other publication AU 7081094 950206.
COMMENT Location/Qualifiers

FEATURES
source 1..35
Location/Qualifiers
/organism="unidentified"
/db_xref="taxon:32644"

BASE COUNT 5 a 5 c 5 g 20 t
ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;

Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 9

LOCUS A46467 35 bp DNA PAT 07-MAR-1997

DEFINITION Sequence 5 from Patent WO9526402.

ACCESSION A46467

NID 92300644

VERSION A46467.1 GI:2300644

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 35)

AUTHORS Knox,D.P., Smith,S.K., Smith,W.D., Redmond,D. and Murray,J.

TITLE VACCINES AGAINST HELMINTHIC PARASITES

JOURNAL Patent: WO 9526402-A 5 05-OCT-1995;

COMMENT MALLINCKRODT VETERINARY INC (US)

Other publication ZA 9502370 951215

Other publication CA 2182178 951005

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Other publication AU 1956495 951017.

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
Db 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 11

LOCUS A68608 35 bp DNA PAT 06-MAY-1999

DEFINITION Sequence 8 from Patent WO9749726.

ACCESSION A68608

NID 94759636

VERSION A68608.1 GI:4759636

KEYWORDS

SOURCE

ORGANISM

REFERENCE 1 (bases 1 to 35)

AUTHORS Mele,A., De,S.R., Parente,D. and Colanaghi,M.I.

TITLE RECOMBINANT RIBOSOMAL INHIBITOR PROTEIN (RIP) AND USE AS

JOURNAL IMMUNOCONJUGATE

Patent: WO 9749726-A 31-DEC-1997;

MINISTERO UNI RICERCA SCIEN E (IT)

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

Other publication IT F1960155 19971229.

RESULT 13
 LOCUS 113679 35 bp DNA PAT 08-AUG-1995
 DEFINITION Sequence 35 from patent US 5439820.
 ACCESSION 113679
 NID 9996745
 VERSION 113679.1 GI:996745
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Sammentos, P., De Taxis du Poet, P., Nitti, G. and Scacheri, E.
 TITLE Anti-thrombin polypeptides
 JOURNAL Patent: US 5439820-A 35 08-AUG-1995;
 FEATURES Location/Qualifiers
 source 1..35
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 58
 ||||||||||||||||||||||||||||||||||||
 Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

RESULT 14
 LOCUS 124027 35 bp DNA PAT 21-NOV-1996
 DEFINITION Sequence 9 from patent US 5541110.
 ACCESSION 124027
 NID 91603897
 VERSION 124027.1 GI:1603897
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Siegall, C.B.
 TITLE Cloning and expression of a gene encoding bryodin 1 from Bryonia
 JOURNAL Patent: US 5541110-A 9 30-JUL-1996;
 FEATURES Location/Qualifiers
 source 1..35
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 58
 ||||||||||||||||||||||||||||||||||||
 Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

RESULT 15
 LOCUS 128284 35 bp DNA PAT 30-OCT-1996
 DEFINITION Sequence 13 from patent US 5569830.
 ACCESSION 128284
 NID 91819060
 VERSION 128284.1 GI:1819060
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

Unclassified.
 REFERENCE 1 (bases 1 to 35)
 AUTHORS Bennett, A., Labavitch, J.M., Powell, A. and Stoltz, H.
 TITLE Plant inhibitors of fungal polygalacturonases and their use to
 JOURNAL control fungal disease
 Patent: US 5569830-A 13 29-OCT-1996;
 FEATURES Location/Qualifiers
 source 1..35
 BASE COUNT 5 a 5 c 5 g 20 t
 ORIGIN

Query Match 60.3%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 58
 ||||||||||||||||||||||||||||||||||||
 Db 1 GACTCGAGTCGACATCGATTTTCTTTTCTTTT 35

Search completed: September 13, 1999, 16:19:30
 Job time: 4482 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:21:48 ; Search time 133.08 Seconds
(without alignments)
109.041 Million cell updates/sec

File: US-09-325-095-1516.SEQ

Sequence: 1 AATTCACACACTGGCATGCC.....CGATTTTTCCTTTTTCCTTTT 58

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	60.3	35	1	036668
2	35	60.3	35	1	052503
3	35	60.3	35	1	059015
4	35	60.3	35	1	057021
5	35	60.3	35	1	071021
6	35	60.3	35	1	063873
7	35	60.3	35	1	065594
8	35	60.3	35	1	078807
9	35	60.3	35	1	084786
10	35	60.3	35	1	090167
11	35	60.3	35	1	089878
12	35	60.3	35	1	0705321
13	35	60.3	35	1	094244
14	35	60.3	35	1	110696
15	35	60.3	901	1	110698
16	35	60.3	35	1	116211
17	35	60.3	35	1	110276
18	35	60.3	35	1	103701
19	35	60.3	35	1	111753
20	35	60.3	37	1	074002
21	35	60.3	35	1	135802
22	35	60.3	35	1	128797
23	35	60.3	35	1	142690
24	35	60.3	35	1	142699
25	35	60.3	35	1	143345
26	35	60.3	35	1	149441
27	35	60.3	35	1	136629
28	35	60.3	35	1	161267
29	35	60.3	35	1	147688
30	35	60.3	35	1	172469
31	35	60.3	40	1	176809
32	35	60.3	35	1	173128
33	35	60.3	35	1	179144
34	35	60.3	40	1	188344
35	35	60.3	35	1	109259
36	35	60.3	35	1	102554
37	35	60.3	36	1	125542
38	35	60.3	35	1	128092
39	35	60.3	35	1	153244
40	35	60.3	35	1	156016
41	35	60.3	35	1	154046
42	35	60.3	35	1	164425
43	35	60.3	35	1	124713

ALIGNMENTS

C	44	33	56.9	744	1	030669
C	45	33	56.9	1193	1	030670

RESULT	1
ID	036668
AC	036668: standard; cDNA; 35 BP.
DT	09-JUN-1993 (first entry)
DE	PPO adaptor primer #1.
KM	Polyphenol oxidase; PPO; catalytic; browning; fruit; plastid; vacuole;
KW	transit peptide; recombinant plasmid; PCR; primer; amplify; broad bean;
KW	potato; polyphenol oxidase; reaction; ss.
OS	Synthetic.
PN	MO302195-A.
PD	04-FEB-1993.
PF	16-JUL-1992; AU0356.
PR	17-JUL-1991; AU-007248.
PA	(CSIR) COMMONWEALTH SCI & IND RES ORG.
PI	Dry IB, Robinson SP;
DR	WPI: 93-058792/07.
PT	DNA encoding polyphenol oxidase polypeptide or fragment - useful
PT	for modifying the oxidase activity in fruit and vegetables to
PS	decrease or enhance browning
PS	Claim 15: Page 23: 44pp: English.
CC	The sequences given in 036668-69 are adaptors which were used in the
CC	cloning of the polyphenol oxidase (PPO) enzyme genes from various
CC	plants. The PPO genes were isolated, and recombinant plasmids for
CC	transformation of plant cells were produced by PCR using the primers
CC	given in 036678-78. PPO is thought to be the predominant catalyst in
CC	browning of fruit caused by injury or damage. PPO is localised in the
CC	plastids of plant cells whereas the phenolic substrates of the enzyme
CC	are stored in the plant cell vacuole. This compartmentation prevents
CC	the browning reaction from occurring unless the plant cells are damaged
CC	and the enzyme and the substrate are mixed. The PPO gene sequences
CC	could be used to construct synthetic genes which may be used to
CC	transform plants to decrease expression of the enzyme gene. In some
CC	instances, eg. coffee, tea, black olives etc., it is desirable to
CC	increase the level of PPO to produce desired levels of browning or
CC	changes in flavour compounds. The grapevine PPO gene codes for an
CC	additional 103 amino acids upstream of the N-terminus of the mature
CC	protein. This region has the properties of a chloroplast transit
CC	peptide and is most likely responsible for targeting of the protein
CC	to be imported into the chloroplast and processed to produce mature
CC	PPO. Transformation of plants with this gene may therefore result
CC	in correct targeting and maturation of the grapevine PPO in other
CC	species and result in accumulation of active grapevine PPO enzyme in
CC	the plastids of these tissues.
CC	Sequence 35 BP: 5 A: 5 C: 5 G: 20 T:

Query Match	60.3%; Score 35; DB 1; Length 35;
Best Local Similarity	100.0%; Pred. No. 0.0013;
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

07	24	GACGAGTCGACGATTTTTCCTTTTTCCTTTT 58
DB	1	GACGAGTCGACGATTTTTCCTTTTTCCTTTT 35

RESULT	2
ID	052503
AC	052503: standard; DNA; 35 BP.
DT	31-MAY-1994 (first entry)
DE	Helminth aminopeptidase H11-1 adaptor primer.
KW	Helminth aminopeptidase; H11-3; H11-2; H11-1; integral membrane;
KW	alpha-amino acyl peptide hydrolase (microsome); multigene family;
KW	antigen; vaccine; parasite; human; virus; microbe; PCR; primer;

KW polymerase chain reaction; amplify; Haemonchus contortus; ss.
 OS Synthetic.
 PN WO9323542-A.
 PD 25-NOV-1993.
 PF 07-MAY-1993: G00943.
 PR 08-MAY-1993: GB-009993.
 RA (AGRI-) AGRIC & FOOD RES COUNCIL.
 PI Graham M, Knox DP, Munn EA, Newton SE, Oliver JJ,
 PI Smith TS;
 DR WPI: 93-36574/48.
 PT New DNA encoding amino-peptidase from Haemonchus contortus and
 PT derived peptide(s) - useful in protective vaccines, for new
 PT vectors, transformed cells and oligosaccharide, for incorporation
 PT in virus or microbe
 PS Example: Page 32: 137pp; English.
 CC The sequences given in 052503-11 are primers which were used to
 CC amplify cDNA clones of the helminth aminopeptidase genes H11-1, -2
 CC and -3. The amplified sequences encode H11D which is a protein
 CC doublet which shows homology to a family of integral membrane amino-
 CC peptidases. The differences between the amplified clones can be
 CC attributed to different mRNAs of the multigene family, and also to
 CC different variants of the H11D-encoding sequence being present at
 CC different stages of the life cycle, or in strains differing in
 CC geographical origin. Antigenic fragments of the aminopeptidases
 CC encoded by the H11 genes may be used in vaccines to stimulate immune
 CC response against helminth parasites in humans or other animals. These
 CC DNA sequences may be incorporated into a virus or microbe and used in
 CC a similar manner.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTT TTTT TTTT 58
 Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 3
 OS9015
 ID 059015 standard; DNA: 35 BP.
 AC 059015.
 DT 12-APR-1994 (first entry)
 DE RACE PCR adaptor-dt.
 KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification; p110.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993: G00761.
 PR 13-APR-1993: GB-008135.
 RA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G,
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 39; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
 CC was performed by synthesizing first strand cDNA from random hexamers
 CC on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
 CC dt as primers. Products were fractionated using an agarose gel. The
 CC DNA was isolated from the gel and subjected to PCR using oligo 2280 and
 CC adaptor as primers. A product of 350 bp was further sequenced.
 CC See also 051155-6, 059012-23 and 057532-3.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTT TTTT TTTT 58
 Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 4
 OS7021
 ID 057021 standard; DNA: 35 BP.
 AC 057021.
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase adaptor dt primer.
 KW 110 kD catalytic subunit; phosphatidylinositol 3-kinase;
 KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KW ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993: G01651.
 PR 05-AUG-1993: GB-016654.
 RA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 DR WPI: 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 15; 71pp; English.
 CC The sequences given in 057020-23 are primers which were used in the
 CC isolation of the phosphatidylinositol (Ptdins) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schizosaccharomyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the ptdins
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the ptdins is induced. Ptdins activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.0013;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTT TTTT TTTT 58
 Db 1 GACTCGAGTCGACATCGATTTT TTTT TTTT 35

RESULT 5
 OS71021
 ID 071021 standard; DNA: 35 BP.
 AC 071021.
 DT 26-SEP-1994 (first entry)
 DE Primer for amplifying peach polygalacturonase coding sequence.
 KW Ripening; melting; pectins; breeding; Prunus; Rosaceae;
 KW endopolygalacturonase; EndoPG; polygalacturonase; PG; ss.
 OS Synthetic.
 PN WO9405795-A.
 PD 17-MAR-1994.
 PF 27-AUG-1993: AU0442.
 PR 27-AUG-1993: AU-004370.
 RA (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA (PACI-) PACIFIC SEEDS PVT LTD.
 PI Brady CJ, Lee E, Lester DR, Orr G, Speirs J;

DR WPI: 94-101199/12.
PT DNA encoding peach polygalacturonase (Pg) and production of
PT recombinant peach Pg in rosaceae plants - for production of
PT plants with 'melting' or 'non-melting' phenotype
PS Disclosure: Page 15; 51pp; English.
CC During early ripening of peaches, tissue firmness decreases
CC slowly and progressively. Towards the end of ripening, loss of
CC tissue firmness is rapid. This second stage of softening is called
CC the 'melting' stage. Fruit of peach varieties used for canning do
CC not have a 'melting' phase of softening. Ripen fruit remain
CC relatively firm and maintain their shape throughout processing.
CC Fruit of 'melting' varieties show an increase in activity of
CC endopolygalacturonase (EndoPG) during ripening. Polygalacturonase
CC (Pc) is believed to contribute to fruit softening through its action
CC on intercellular and cell wall pectins. This primer was used to
CC synthesise a first strand cDNA molecule for further amplification
CC by PCR. See 056765-67.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 6
Q63873
ID Q63873 standard; cDNA; 35 BP.
AC Q63873;
DT 04-NOV-1994 (first entry)
DE T cell protein CD4 adaptor-drl1 primer.
KW Cat: feline immunodeficiency virus; CD4 positive T lymphocyte;
KW cellular receptor; T cell protein; polymerase chain reaction; PCR; ss.
OS Synthetic.
PN FR2696188-A.
PD 01-APR-1994.
PF 25-SEP-1992: 011496.
PR 25-SEP-1992: FR-011496.
PA (UYPA-) UNIV CURIE PARIS VI P & M.
PI De Parseval A, Klatzmann D, Salmon P, Sonigo P;
DR WPI: 94-128285/16.
PT Feline CD4 protein and corresp. DNA - for study of feline
PT immunodeficiency virus and for therapy
PS Disclosure: Page 5; 25pp; French.
CC The primers Q63871 and Q63872 were used to obtain cDNA coding for
CC feline CD4 protein by polymerase chain reaction. For rapid
CC amplification of the ends of the cDNA, a hybrid adaptor-oligo(dt)
CC primer was also used. Cat lymph node cDNA was amplified and the
CC full-length CD4 coding sequence Q63870 was obtained from several
CC partial clones.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 7
Q63594
ID Q63594 standard; DNA; 35 BP.
AC Q63594;
DT 12-JAN-1995 (first entry)
DE HER4 primer XSC17.
KW Human epidermal growth factor; receptor; HER; EGF; tyrosine kinase;

KW cancer; neuronal tissue; muscle tissue; neoplasm;
KW carcinoma; primer; probe; PCR; ss.
OS Synthetic.
PN EP-599274-A.
PD 01-JUN-1994.
PR 23-NOV-1993; 118837.
PR 24-NOV-1992; US-981165.
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Culouscou J, Ploymann GD, Shoyab M;
DR WPI: 94-169599/21.
PT New recombinant nucleic acid expressing HER4 - a new receptor
PT tyrosine kinase expressed in some cancer cells, and related
PT vectors, antibodies, ligands etc, for diagnosis and treatment of
PT cancers
PS Example: Page 16; 104pp; English.
CC The primers given in 065594-6 were used in the semi-quantitative
CC PCR detection of HER4.
CC HER4 is the fourth member of the EGFR-family of tyrosine kinases and
CC is expressed in some human cancers and in some tissues of neuronal
CC or muscle origin. HER4 polynucleotides, opt. labelled, are useful
CC in assays (e.g. of HER4 mRNA to detect certain neoplasms, esp.
CC breast carcinoma) and as primers in PCR or as probes.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||
DB 1 GACTCGAGTCGACATCGATTTTGTGTTT 35

RESULT 8
Q78807
ID Q78807 standard; cDNA; 35 BP.
AC Q78807;
DT 12-JUN-1995 (first entry)
DE TCR alpha-chain dr17 adaptor.
KW Allergen: immunotherapy; T-cell receptor; TCR; CDR3;
KW complementarity determining region-3; immunosuppressant;
KW antibody engineering; suppressor T-lymphocyte; Ts; PCR;
KW polymerase chain reaction; amplification; adaptor; ss.
OS Synthetic.
PN WO9425489-A.
PD 10-NOV-1994.
PF 26-APR-1994; CA0228.
PR 26-APR-1993; GB-008581.
PA (UYMA-) UNIV MANITOBA.
PI Mohapatra SS, Sehon AH;
DR WPI: 94-358193/44.
PT Synthetic peptide(s) for effecting immuno-therapy - have an amino
PT acid sequence comprising a portion of the CDR3 region of a T-cell
PT receptor
PS Disclosure: Page 43; 78pp; English.
CC Antibody production against allergenic antigen (Ag) is specifically
CC suppressed by treatment with the junctional segment of the alpha
CC and/or beta chain of the Ag receptor of suppressor T (Ts) cells
CC (TcR) induced by tolerogenic Ag-PEG conjugates. cDNAs encoding
CC the alpha and beta chains of TCRs of cloned Ts cells specific for
CC ovalbumin and human monoclonal (myeloma) IgG were produced by PCR
CC using the primers given in 078798-805 and the adaptors given in
CC 078806-07.
SQ Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTTGTGTTT 58
|||||

```

Db      1 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT
RESULT 9
ID      084786 standard; DNA: 35 BP.
AC      084786;
DE      17-AUG-1995 (first entry)
DE      PTK22 reverse-transcription primer.
KW      Protein-tyrosine-kinase; PTK; discoidin domain receptor; cancer;
KW      breast tumor; mamma carcinoma; diagnosis; prognosis; therapy; ss.
OS      Synthetic.
PN      MO9502187-A.
PD      19-JAN-1995.
PF      08-JUL-1994; G01480.
PR      09-JUL-1993; GB-014271.
PA      (CANC-) CANCER RES INST.
PA      (WELL) WELLCOME FOUND LTD.
PI      Barker KT, Crompton MR, Gusterson BA, Martindale JE;
PI      Mitchell PJ, Page MJ, Spence P;
PI      WPI: 95-066991/09.
PT      Method for screening substances, using protein tyrosine kinase -
PT      for potential utility as therapeutic agents for cancer
PS      disclosure; Page 34; 51pp; English.
CC      cDNA derived from tumor metastatic tissue was amplified using
CC      primers (given in 084783-84) based on sequences (R71101, R71103)
CC      associated with protein-tyrosine-kinases (PTK). Novel PTK22 was
CC      identified in an isolated subclone. The 3' sequence of PTK22 was
CC      obtained by reverse transcription (using the primer of 084786) and
CC      PCR amplification (primers 084787-88) of RNA of human breast
CC      carcinoma cell line MDA MG 468. The partial DNA sequence of PTK22
CC      is given in 084782.
SQ      Sequence 35 BP: 5 A; 5 C; 5 G; 20 T;

Query Match
Best Local Similarity 100.0%; Score 35; DB 1; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      24 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT 58
      |||||||
Db      1 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT 35

RESULT 10
ID      090167 standard; DNA: 35 BP.
AC      090167;
DE      01-NOV-1995 (first entry)
DE      Pollen allergen gene primer 8.
KW      Japanese cedar; pollen; allergy; therapy; diagnostic;
KW      desensitizer; Cryptomeria japonica; polymerase chain reaction;
KW      PCR; primer; ss.
OS      Synthetic.
PN      EP-655500-A.
PD      31-MAY-1995.
PF      03-NOV-1994; 308117.
PR      05-NOV-1993; JP-299151.
PR      20-DEC-1993; JP-344596.
PR      27-DEC-1993; JP-346814.
PA      (HAYB) HAYASHIBARA SEIBUTSU KAGAKU.
PI      Kunitomo M, Namba M, Toriyoue K;
PI      WPI: 95-195588/26.
PT      New Japanese cedar pollen allergen polypeptide - and DNA coding for it,
PT      useful for treatment and diagnosis of cedar pollen allergy
PS      disclosure; Page 11; 41pp; English.
CC      The primers given in 090166-69, based on a portion of cDNA clone
CC      SC09 or on previously isolated sequences, were used for the PCR
CC      amplification of cDNA derived from Japanese cedar pollen, to obtain
CC      clone SC50 bearing a partial sequence (nt 1-198 of the sequence
CC      given in 090156) of the pollen allergen gene.
SQ      Sequence 35 BP: 5 A; 5 C; 5 G; 20 T;

```

```

Query Match
Best Local Similarity 100.0%; Score 35; DB 1; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      24 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT 58
      |||||||
Db      1 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT 35

RESULT 11
ID      089878 standard; DNA: 35 BP.
AC      089878;
DE      31-JAN-1997 (Revised)
DE      25-OCT-1995 (first entry)
DE      PCR primer for cloning bryodin-2.
KW      Ribosome inactivating protein; cell death; proliferation; tumour;
KW      cancer; virus; malaria; trypanosomiasis; inflammation; autoimmune;
KW      disease; purging; bone marrow; ss.
OS      Synthetic.
PN      MO9511977-A.
PD      04-MAY-1995.
PF      25-OCT-1994; U12382.
PR      25-OCT-1993; US-141891.
PR      20-OCT-1994; US-324301.
PA      (BRIT) BRISTOL-MYERS SQUIBB CO.
PI      Gawlak SL, Margardt H, Siegal CB;
PI      WPI: 95-178870/23.
PT      New ribosome inactivating protein bryodin-2 and related
PT      conjugates - for killing target cells, e.g. in treatment of
PT      cancer, also related nucleic acid, vectors and transformed cells
PS      disclosure; Page 43; 81pp; English.
CC      The sequence is that of a PCR primer used to isolate a ribosome
CC      inactivating protein termed bryodin-2, from Bryonia dioica. Conjugates
CC      of the protein with a ligand may be used to kill target cells (i.e. those
CC      to which the ligand component binds specifically) or to inhibit
CC      proliferation of tumor cells. Typical applications include the
CC      treatment of cancer, viral infections, malaria, trypanosomiasis,
CC      inflammatory or autoimmune diseases, including in vitro purging of e.g.
CC      bone marrow.
CC      See also 089874-86.
CC      updated 31-JAN-1997 due to incorrect cross reference numbers.
SQ      Sequence 35 BP: 5 A; 5 C; 5 G; 20 T;

Query Match
Best Local Similarity 100.0%; Score 35; DB 1; Length 35;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      24 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT 58
      |||||||
Db      1 GACTCGAGTCGACATCGATTTT TTTT TTTT TTTT 35

RESULT 12
ID      T05321 standard; DNA: 35 BP.
AC      T05321;
DE      13-APR-1996 (first entry)
DE      Oligonucleotide probe for human prostacyclin synthase.
KW      DNA probe; prostacyclin-synthase; PCR; polymerase chain reaction;
KW      diagnosis; gene therapy; ss.
OS      Synthetic.
PN      MO9530013-A1.
PD      09-NOV-1995.
PF      27-APR-1995; J00838.
PR      28-APR-1994; JP-114316.
PA      (TANA/) TANABE T.
PI      Tanabe T;
PI      WPI: 95-393084/50.
PT      Human prostacyclin synthase and DNA encoding it - useful in the

```

PT Investigation and treatment of diseases characterised by reduced prostaglandin 12 production.

PS Disclosure: Page 34; 71pp; Japanese.

CC DNA primers (T05317-20; T05322; T05326-27) are used to screen human genomic lung cell line W138 and human arterial endothelial cell CC cDNA libraries for the isolation of a prostacyclin synthase (PGIS) coding sequence (see T05316). DNA probes (T05321 and T05323) are used in the construction of plasmid pHPGIS1, encoding the complete PGIS sequence. This plasmid was used to transfect human 293 cells for CC PGIS peptide expression. DNA encoding human PGIS, vectors containing CC it, and PGIS itself, may be administered to patients to increase prostaglandin 12 (PGI2) production to treat diseases characterized CC by reduced PGI2 levels or by an imbalance between PGI2 and thromboxane A2 levels, such as circulatory diseases (thrombosis, CC angina pectoris, arteriosclerosis, myocardial infarction). The CC DNA and protein are also useful in disease diagnosis.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTTCCTTTT 35

RESULT 13

ID 094244
AC 094244; standard; DNA; 35 BP.

DT 1A-MAY-1996 (first entry)
KW Cysteine proteinase active site antisense polyT primer.
NM Primer: amplify; polymerase chain reaction; PCR; Haemonchus contortus;
KM cysteine proteinase; DM.2; DM.3; DM.4; human; DM.4a; DM.5; antigen;
KW vaccine; helminth; parasite; DM.1; ruminant; gastro-intestinal tract; ss.
OS Synthetic.

PN W09526402-A1.
PD 05-OCT-1995.
PF 24-MAR-1995; G00665.
PR 25-MAR-1994; GB-005990.
PR 25-MAR-1994; GB-005925.
PA (MLCV) MALLINCKRODT VETERINARY INC.
PI Knox DP, Murray J, Redmond D, Smith SK, Smith WD;
DR WPI; 95-351322/45.
PT Protective helminth parasite antigen - used in vaccine directed PT against parasitic nematodes of mammalian gastro-intestinal tract e.g. Haemonchus contortus

PS Example 16; Fig 15; 79pp; English.

CC The sequences given in 094240-45 are primers which are based on CC previously published sequences derived from the canonical Haemonchus contortus cysteine proteinase molecule. These primers were used in CC the cloning of cDNA fragments from the cysteine proteinase gene, CC such as DM.1, DM.2, DM.3, DM.4, DM.4a and DM.5 (see also 094246-51). CC The amplified fragments may be expressed in a recombinant cell for CC the production of antigens. These antigens may be used in the CC preparation of a vaccine against helminth parasites in a human or CC non-human animal.

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTTCCTTTT 35

RESULT 14

T10696

ID T10696 standard; DNA; 35 BP.

AC T10696;

DT 26-MAY-1996 (first entry)

DE Trypsin-like enzyme oligonucleotide TRX-10.

KW Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW protease; primer; PCR; amplification; ss.
OS Synthetic.

PN A09527248-A.
PD 08-FEB-1996.
PF 31-JUL-1995; 027248.
PR 29-JUL-1994; JP-178607.
PA (TEIJ) TEIJIN LTD.
PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;
PI Yamaoka K, Yasuoka S;
DR WPI; 96-117356/13.
PT Nucleic acid sequence encoding trypsin-like enzyme - which digests PT fibrinogen, used as expectorant in treatment of respiratory PT diseases, e.g. bronchial asthma

PS Example 9; Page 44; 65pp; English.

CC Example 9 describes the cloning of cDNA region encoding CC trypsin-like enzyme isolated from cough phlegm.
CC trypsin-like enzyme isolated from cough phlegm.
CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8 (T10695), TRX-10 (T10696) and TRX-11 (T10697).

CC TRX-1 corresponds to from A1 to A23 of the sequence given in CC T10693, which is part of a cDNA encoding the trypsin-like CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.
CC TRX-10 is capable of annealing to the 3'-terminus of poly(A)+ RNA.
CC TRX-11 is identical to the 5'-terminus side 19 residues of TRX-10.
CC After amplification, plasmid p19-33 was obtained. p19-33 encodes CC part of the N-terminus amino acid sequence 20 residues of the CC trypsin-like enzyme isolated from the cough phlegm (see T10698).

SO Sequence 35 BP; 5 A; 5 C; 5 G; 20 T;

Query Match 60.3%; Score 35; DB 1; Length 35;
Best Local Similarity 100.0%; Pred. No. 0.0013;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 24 GACTCGAGTCGACATCGATTTTTCCTTTT 58
DB 1 GACTCGAGTCGACATCGATTTTTCCTTTT 35

RESULT 15

ID T10698/c

ID T10698 standard; cDNA to mRNA; 901 BP.

AC T10698;

DT 26-MAY-1996 (first entry)

DE Trypsin-like enzyme N-terminal DNA fragment from p19-33.

KW Trypsin; fibrinogen; thrombin; expectorant; respiratory disease;
KW asthma; VIP; vasoactive intestinal peptide; influenza virus;
KW protease; primer; PCR; amplification; ds.
OS Homo sapiens.

PN A09527248-A.
PD 08-FEB-1996.
PF 31-JUL-1995; 027248.
PR 29-JUL-1994; JP-178607.
PA (TEIJ) TEIJIN LTD.
PI Masuda K, Ogawa H, Suga T, Sugimoto Y, Takagi K;
PI Yamaoka K, Yasuoka S;
DR WPI; 96-117356/13.
PT Nucleic acid sequence encoding trypsin-like enzyme - which digests PT fibrinogen, used as expectorant in treatment of respiratory PT diseases, e.g. bronchial asthma

PS Example 9; Page 44-45; 65pp; English.

CC Example 9 describes the cloning of cDNA region encoding CC trypsin-like enzyme isolated from cough phlegm.
CC Four oligonucleotides are used for PCR: TRX-1 (T10694), TRX-8 (T10695), TRX-10 (T10696) and TRX-11 (T10697).
CC TRX-1 corresponds to from A1 to A23 of the sequence given in CC T10693, which is part of a cDNA encoding the trypsin-like CC enzyme. TRX-8 corresponds to from G16 to T40 of T10693.

CC TRY-10 is capable of annealing to the 3'-terminus of poly(A)+ RNA.
CC TRY-11 is identical to the 5'-terminus side 19 residues of TRY-10.
CC After amplification, plasmid p19-33 was obtained. p19-33 encodes
CC part of the N-terminus amino acid sequence 20 residues of the
CC trypsin-like enzyme isolated from the cough phlegm (see T10698).
SQ Sequence 901 BP: 270 A; 185 C; 205 G; 241 T;

Query Match 60.3%; Score 35; DB 1; Length 901;
Best Local Similarity 100.0%; Pred. No. 0.0024;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 24 GACTCGAGTCGACATCGATTTT
Db 901 GACTCGAGTCGACATCGATTTT 867

Search completed: September 13, 1999, 16:21:48
Job time: 4008 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:13:49 ; Search time 849.94 Seconds

(Without alignments)
134,606 Million cell updates/sec

Title: US-09-325-095-1516.SEQ

Perfect score: 58
Sequence: 1 AATTCACACACTGGCATGCC.....CGATTTCATTTTTCATTTT 58

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	38	65.5	337	32	AA340759	AA340759 EST46034
2	28.2	48.6	241	44	A1270969	A1270969 gw52f12.x
3	27.8	47.9	541	47	A1479049	A1479049 tm03e03.x
4	27.8	47.9	474	48	A1568396	A1568396 th13g06.x
5	27	46.6	477	43	A122956	A122956 EST22651
6	26.8	46.2	397	25	N89796	N89796 zb34b04..s1
7	26.8	46.2	498	39	AA833826	AA833826 od62c05.s
8	26.8	46.2	425	39	AA902569	AA902569 o159d09.s
9	26.8	46.2	765	41	A1026930	A1026930 ov82e09.x
10	26.8	46.2	456	42	A1079505	A1079505 oy95f01.x
11	26.8	46.2	511	44	A1242489	A1242489 q116f01.x
12	26.8	46.2	723	45	A1378815	A1378815 lc67d11.x
13	26.8	46.2	527	46	A1445584	A1445584 t326a11.x
14	26.8	46.2	600	48	A1568176	A1568176 tn66d10.x
15	26.6	45.9	207	50	AV025932	AV025932 AV025932
16	26.4	45.5	452	25	N50597	N50597 yf89e06.r1
17	26.4	45.5	435	28	C18175	C18175 C18175 Huma
18	26.4	45.5	399	34	AA509647	AA509647 vhs1g12.r
19	26.4	45.5	499	37	AA675277	AA675277 v165f10.s
20	26.4	45.5	390	44	A1255155	A1255155 u132a11.y
21	26.4	45.5	616	51	A1725226	A1725226 1245 PRIF
22	26	44.8	478	26	W17126	W17126 zbl7f05.r1
23	26	44.8	492	28	AA11945	AA11945 z88a08.r
24	26	44.8	338	29	AA16347	AA16347 zp04g09.r
25	26	44.8	180	36	C61881	C61881 C61881 Yuj1
26	26	44.8	352	26	A1640742	A1640742 wai15a02.x
27	25.8	44.5	698	26	W20094	W20094 zb39d09..r1
28	25.8	44.5	570	29	AA145612	AA145612 ms07h10.r
29	25.8	44.5	357	37	AA668836	AA668836 ab91g02.s
30	25.8	44.5	656	44	A1322793	A1322793 ms07h10.y
31	25.6	44.1	678	48	A1295887	A1295887 lrp09649.5
32	25.6	44.1	287	48	A1545931	A1545931 fb66b05.y
33	25.6	44.1	534	48	A1557358	A1557358 PT2.1.6.B
34	25.6	44.1	1100	51	AU050243	AU050243 AU050243
35	25.4	43.8	317	41	A1002468	A1002468 oq83h06.s
36	25.4	43.8	349	44	A1305227	A1305227 qm13f04.x
37	25.4	43.8	111	44	A1312326	A1312326 ta78e03.x
38	25.4	43.8	380	45	A1348883	A1348883 lb05c11.x
39	25.4	43.8	909	48	A1568751	A1568751 th15d09.x
40	25.4	43.8	408	48	AU036448	AU036448 AU036448
41	25.4	43.8	466	51	A1717729	A1717729 ut-r-10-a
42	25.4	43.8	458	51	A1733464	A1733464 oq83h06.x
43	25.2	43.4	392	20	D31233	D31233 HUML1301 Hu
44	25.2	43.4	244	20	T02622	T02622 0276C3 cbps
45	25.2	43.4	567	31	AA291389	AA291389 zt38a03.s

ALIGNMENTS

RESULT 1
AA340759 337 bp mRNA EST 21-APR-1997
LOCUS EST46034 Fetal Kidney II Homo sapiens cDNA 5' end similar to
DEFINITION similar to protein phosphatase 2A, beta, 55 kDa, mRNA sequence.
ACCESSION AA340759
GI992998
VERSION AA340759.1 GI:1992998

KEYWORDS EST.
 human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 337)
 AUTHORS Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,
 White, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,
 Clifton, A., Sutton, G., Blake, J.A., Brandon, R.C., Man, M.L., C.
 Clifton, R.A., Cline, T.R., Colton, M.D., Earle-Hughes, J., Fine, L.D.,
 Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geohagen, N.S.,
 Glodok, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,
 Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaro, S.M., Merrick, J.M.,
 Moreno-Altamirano, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,
 Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,
 Smith, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,
 Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,
 Dimke, D., Feng, D.-F., Ferris, A., Fischer, C., Hastings, G.A.,
 He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,
 Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H.,
 Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,
 Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,
 Fraser, C.M. and Venter, J.C.
 Initial assessment of human gene diversity and expression patterns
 based upon 83 million nucleotides of cDNA sequence
 Nature 377 (6547 Suppl.), 3-174 (1995)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1406936.
 Other ESTs: EST46035 THC125647
 Contact: Kerlavage, AR
 *Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: arkerlav@tigr.org
 For clone availability, additional sequence and expression
 information related to this EST, please check the TIGR Human Gene
 Index (<http://www.tigr.org/cdb/hgi/hgi.html>)
 Seq primer: M13-21.
 FEATURES
 source
 Location/Qualifiers
 1..337
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):142407"
 /db_xref="taxon:9606"
 /clone_lib="Fetal Kidney II"
 /dev_stage="fetus"
 /note="Organ: Kidney; Vector: pBluescript KS-; Site_1:
 XhoI; Site_2: EcoRI"
 BASE COUNT 63 a 88 c 95 g 88 t 3 others
 ORIGIN
 Query Match 65.5%; Score 38; DB 32; Length 337;
 Best Local Similarity 100.0%; Pred. No. 0.06;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 21 GATGACTGAGTCGACATCGATTTTGTGTGTGT 58
 Db 1 GATGACTGAGTCGACATCGATTTTGTGTGTGT 38
 RESULT 2
 LOCUS A1270969 241 bp mRNA EST 21-DEC-1998
 DEFINITION gw5212.1 x1 NCI-CGAP_Br13 Homo sapiens cDNA clone IMAGE:1994735 3',
 mRNA sequence.
 ACCESSION A1270969
 NID 93890136
 VERSION A1270969.1 GI:3890136
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 241)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Aug 21, 1998 this sequence version replaced.
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 CDNA Library Preparation: David B. Krizman, Ph.D.
 CDNA Library Arrayed by: I.M.A.G.E. Consortium, LNL
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/dbp/image/image.html
 Insert Length: 299 Std Error: 0.00
 Seq primer: -40UP from G1bco.
 FEATURES
 source
 Location/Qualifiers
 1..241
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1994735"
 /clone_lib="NCI-CGAP_Br13"
 /sex="female"
 /tissue_type="Breast carcinoma in situ"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: breast; Vector: PAMPI: mRNA made from breast
 carcinoma, cDNA made by oligo-dT priming. Directionally
 cloned. Size-selected on agarose gel, average insert
 size 500 bp. Primary library, non-amplified."
 BASE COUNT 77 a 63 c 62 g 38 t 1 others
 ORIGIN
 Query Match 48.6%; Score 28.2; DB 44; Length 241;
 Best Local Similarity 72.0%; Pred. No. 37;
 Matches 36; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
 QY 9 CATGCGATGCCGATGACATCGATCGATTTTGTGTGTGT 58
 Db 235 CATGCGATGCCGATGACATCGATCGATTTTGTGTGTGT 186
 RESULT 3
 LOCUS A1479049 541 bp mRNA EST 14-APR-1999
 DEFINITION tm30e03.x1 NCI-CGAP_CL11 Homo sapiens cDNA clone IMAGE:2158108 3',
 similar to contains Alu repetitive element; mRNA sequence.
 ACCESSION A1479049
 NID 94372217
 VERSION A1479049.1 GI:4372217
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 541)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3187992.
 CONTACT: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Ash Alizadeh, John Byrd, M.D., Mike Grever,
 M.D., Louis M. Staudt, M.D., Ph.D.

Best Local Similarity 68.5%; Pred. No. 81;
Matches 37; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

0v 5 CACACACTGGCATGCCGATGACTCGAGTCGACATCGATTTTTTTTTTTTT 58
 ||||| | | | | | | | | | | | | | | | |
Db 257 CACACACTGGCTGTGTGAGAAGCTCAAGGCTAATTTTTTATCCCTTTTTTTTTTT 20

RESULT	8
AA902569/c	
...CCTG	
DEFINITION	
AA902569	425 bp
	mRNA
oj59d09.s1 NCI_CGAP_kid3 Homo sapiens cDNA clone IMAGE:1507609	EST
	09-JUN-1998
	3'

ACCESSION	AA902569	
NID	93037692	
VERSION	AA902569.1	GI:3037692
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
Eutheria; Primates; Catarrhini; Hominidi; Homo.
1 (bases 1 to 425)
REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS National Cancer Institute Cancer Genome Initiative
TITLE National Cancer Institute Cancer Genome Initiative Project (CCG)

JOURNAL Unpublished (1997)
COMMENT On Jan 9, 1998 this sequence version replaced g1:936995

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution Information can be found through the I.M.A.G.E. Consortium/LLNL at: www.dlo.llnl.gov/bdrip/image/image.html

Trace considered overall poor quality
Insert Length: 1525 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 1.

```

/organism="homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1502609"
/clone_lib="NCI_CGAP_Kid3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I ; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

```

Query Match	46.2%	Score 26.8	DB 39	length 425
Best Local Similarity	68.5%	Pred. No. 83		
Matches 37	Conservative	0	Mismatches 17	Indels 0
				Gaps 0

QY 5 CACACACTGGCATGCCGATGCTCGAGTGCACATCGATTTTTTTTTTTTTTTT 58
 ||||| | | ||| | | | | | | |
Db 92 CACACACTGGCTGGTGAGACTCAAGGCTAATTTTTCCTTTTTTTTTTTTTTT 39

RESULT	9
LOCUS	A1026930/c
DEFINITION	A1026930 765 bp mRNA EST 27-AUG-1998
	ov82e09.xl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1643848

ACCESSION	AI026930
NID	g3244446
VERSION	AI026930.1
KEYWORDS	EST

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 765)
NOT-RCAP <http://www.ncbi.nlm.nih.gov/ncbiinfo>

JOURNAL Unpublished (1997)
COMMENT On May 8, 1995 this sequence version replaced g1:801259

Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLT at:
www-bio.lnl.gov/dbfp/image/image.html

Insert Length: 2811 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 466.

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1643848"
/clone_1lb="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTACCAATCTGATGGAGCGGCCGCCCAATTTTTTTTTTTTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73D vector. Library
went through one round of normalization to Cot5 and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

Query Match	46.28	Score 26.8	DB 41	Length 765
Best Local Similarity	68.58	Pred. No. 75		
Matches 37	Conservative 0	Mismatches 17	Indels 0	Gaps 0

QY 5 CACACACTGGCATGCCGATGCAGTGCAGATCGATTTTTTTTTTTTT 58
||||| - ||||| - | | |||||
Db 229 CACACACTGGCTGTGAGAAGCTCAAGCTAATTTTTCCTTTT 17

RESULT 10	LOCUS	DEFINITION
A1079505/c	A1079505	456 bp mRNA EST 29-SEP-1998
	09955f01.x1	Soares fetal liver spleen INFIL S1 Homo sapiens CDNA
	clone IMAGE:1673553	3' similar to gb:M87503 TRANSCRIPTIONAL

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:30 ; Search time 329.34 Seconds

(without alignments)
511.801 Million cell updates/sec

Title: US-09-325-095-1718.SEG

Perfect score: 1 TTTAGCTGAGCATTCCTAA.....CCGACTCGAGTCGACATCGA 53

Scoring table:

IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pl4.*
11: gb_pl5.*
12: gb_pl6.*
13: gb_pl7.*
14: gb_pl8.*
15: gb_pl9.*
16: gb_pl10.*
17: gb_pl11.*
18: gb_pl12.*
19: gb_pl13.*
20: gb_pl14.*
21: gb_pl15.*
22: gb_pl16.*
23: gb_pl17.*
24: gb_pl18.*
25: gb_pl19.*
26: gb_pl20.*
27: gb_pl21.*
28: gb_pl22.*
29: gb_pl23.*
30: gb_pl24.*
31: gb_pl25.*
32: gb_pl26.*
33: gb_pl27.*
34: gb_pl28.*
35: gb_pl29.*
36: gb_pl30.*
37: gb_pl31.*
38: gb_pl32.*
39: gb_pl33.*
40: gb_pl34.*
41: gb_pl35.*
42: gb_pl36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	35	66.0	35	5	A37245	A37245 Sequence 14

C	2	26	49.1	3207	3	BOVPHOS3KN	M93252 Bovine phop
C	3	26	49.1	3498	5	A37232	A37232 Sequence 1
C	4	24.4	46.0	3424	10	HSPH13K	229090 H.sapiens m
C	5	24.4	46.0	3207	10	HSPH143	U79143 Human phosp
C	6	24.4	46.0	3207	12	MMUCG279	U03279 Mus musculu
C	7	24	45.3	166924	7	MIATGENB	Y08502 A.thaliana
C	8	24	45.3	2656	7	SCYOR007C	Z74915 S.cerevisia
C	9	24	45.3	75812	8	ATAC007143	AC007143 Arabidops
C	10	24	45.3	30469	8	SCU3491	U34491 Saccharomyc
C	11	23.8	43.4	15532	37	TCU009810	U09810 Katharina t
C	12	23	43.4	9829	7	CHU68040	U68040 Cochlicobol
C	13	22.8	43.0	3452	4	AF001075	AF001075 Gallus ga
C	14	22.8	43.0	213	12	AF155918	AF155918 Sigmodon
C	15	22.8	43.0	213	12	AF155922	AF155922 Sigmodon
C	16	22.8	43.0	3389	17	AF001075	AF001075 Avian sar
C	17	22.8	43.0	39482	36	CET00448	Z35663 Caenorhabd
C	18	22.6	42.6	1702	7	GMU36590	U36590 Glomus mani
C	19	22.6	42.6	155116	34	HSS31H16	AL031664 Homo sapi
C	20	22.4	42.3	4916	36	AF051936	AF051936 Homarus a
C	21	22.2	41.9	138636	10	HS1045021	AL021919 Human DNA
C	22	22.2	41.9	506	36	MIESLR	X79592 E.sclopes
C	23	22	41.5	78874	7	AB018107	AB018107 Arabidops
C	24	22	41.5	152383	10	HSP18108	AJ006995 Homo sapi
C	25	22	41.5	413	14	G35950	G35950 STS h14a133
C	26	22	41.5	157287	35	AC006402	AC006402 Drosophil
C	27	21.8	41.1	1383	1	MCAAT	Z33356 M.capricolu
C	28	21.8	41.1	1418	7	CPLRRNMA	X68365 C.purpurea
C	29	21.8	41.1	687	7	D89919	D89919 Armillaria
C	30	21.8	41.1	590	7	D89924	D89924 Armillaria
C	31	21.8	41.1	27405	10	HSE95B1	Z69042 Human DNA S
C	32	21.8	41.1	35353	36	CEZK520	Z92822 Caenorhabd1
C	33	21.8	41.1	612	37	TCU89777	U89777 Trypanosoma
C	34	21.8	41.1	613	37	TCU89777	Z92822 S.pombe chr
C	35	21.6	40.8	38078	7	SPAC9E9	AC006756 Caenorhab
C	36	21.6	40.8	231305	34	AC006756	Z96047 Caenorhabd1
C	37	21.6	40.8	42734	36	CEY3	AF000191 Caenorhab
C	38	21.6	40.8	32911	36	CELT2306	AF076243 Arabidops
C	39	21.4	40.4	99384	8	AF076243	U13176 Papio anubi
C	40	21.4	40.4	5417	9	BABAPOCID	AC005350 Homo sapi
C	41	21.4	40.4	78661	11	AC005350	L47550 Mus musculu
C	42	21.4	40.4	3213	12	MUSSTAF	AL022276 Caenorhab
C	43	21.4	40.4	110000	34	CEY102F5_0	AC007802 Drosophil
C	44	21.4	40.4	60649	35	AC007802	AC004366 Drosophil
C	45	21.4	40.4	86884	36	AC004366	

ALIGNMENTS

RESULT 1.	LOCUS	SEQUENCE	14	FROM	PATENT	MO9403609.	PAT	05-MAR-1997
A37245	Sequence 14	35 bp	DNA					
A37245	Accession	g2294356						
A37245.1	GI:2294356							

REFERENCE 1 (bases 1 to 35)
Parker,P.J., Goode,N.T., Nurse,P.M. and Waterfield,M.D.
EUCHARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
OR PROTEIN KINASE AND ASSAYS USING THEM
Patent: WO 9403609-A 14 17-FEB-1994.
IMP CANCER RES. TECH (GB)
Other publication JP 8503124T 960409.
Location/Qualifiers
1..35
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 10 a 9 c 4 g 12 t

Query Match 66.0%; Score 35; DB 5; Length 35;
 Best Local Similarity 100.0%; Pred. No. 0.00029;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTTAAGCTTAAGCTTCTAAAGTCACTATCATCC 35
 ||||||||||||||||||||||||||||||||||||
 Db 1 TTTAAGCTTAAGCTTCTAAAGTCACTATCATCC 35

RESULT 2

BOVPHOS3K/c

LOCUS BOVPHOS3K 3207 bp mRNA MAM 19-AUG-1992
 DEFINITION Bovine phosphatidylinositol 3-kinase 110 kDa subunit mRNA, complete cds.

ACCESSION M93252
 NID G163519
 VERSION M93252.1 GI:163519
 KEYWORDS Phosphatidylinositol 3-kinase.
 SOURCE Bos taurus cDNA to mRNA.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos

REFERENCE 1 (bases 1 to 3207)
 Hiles, I.D., Otsu, M., Volinia, S., Fry, M.J., Gout, I., Dhan, R.,
 Panayotou, G., Ruiz-Larrea, F., Thompson, A.S., Totty, N.F.,
 Hsuan, J.J., Courtneidge, S.A., Parker, P.J. and Waterfield, M.D.
 Phosphatidylinositol 3-kinase: Structure and expression of the 110
 kd catalytic subunit
 Cell 70, 419-429 (1992)

JOURNAL MEDLINE
 FEATURES
 source location/Qualifiers
 1..3207
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /cell_line="SGBR-1"
 1..3207
 /EC_number="2.7.1.67"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase"
 /protein_id="AAA30698.1"
 /db_xref="PID:g163520"
 /db_xref="GI:163520"

CDS

translation="MPRRSGELMGHILMPRIIVLECLPNGMIVTLECLREATLIT
 IKHELFKARKYPLHQLODESSYIFVSTQAEAREEFDETRRLCDLRFQPLKVI
 EPVGRREKILNREIGFAIGMPCEDKPEVDPRNLIINCKEAVDLRLDSPH
 SRAMVYPPNVSSELPKHIYKRLDKGQIIYIVIVYSPNNOKKXTLKHHCQVPE
 OYIAEAIKRTSRMLSSBOLKCYLEVOGYIILKVCQDETFLEKPIPSQYKTRSC
 IMLGRPNMLMAKESLSQLPMDCTMPSTSRISTATPYNGETSTKSLMVINSAL
 RIKICATYVNVNIRIDIKIYRTGIIYHGGELCDNVNTQRPVCSNPRNMLNVDIY
 IPDLPRARLCLISCSYGRGAKKEHCPKANGINLFDYDTLVSGKMLNLPVPH
 GLEDLNPIGVTSNPKETPCLEPEPFMSYVYKPPMSYIEEHANVSREGEFSY
 SHAGISNRLARDNELRENDKQLRAICTRDLSETTDEKDFLSHRHVCYTIPEIIP
 KILLSYKNSDEVAQMTCLYKMDPIRPEBAMELDCNTPDPWVRGFAVCELYLT
 DDKLSQYLIQVQVAKTYEQYLDNLVRLKALNORIGHEFFHMLKSEHNNTVSO
 RFGLLSEYCRACGMYLKHLNROYAMEKLINLTDILKOEKDETOKVMKFLVQMR
 RPFMDALOGFSLNPAHQNLNLEECRIMSARPLMLNMPNDIMSELIFQNNH
 IIFKNGDRLQDMFLTIQIRIMENIMONOGDLMLPYGCLSIDCVGLIEVRSHT
 IMOICGKGLGALQPNSTLHOMLKXNKGEIYDAADLPTRSCAGCVATFTLLGIG
 DRHNSNIVKDDGOLFHDGFHFLDHKKKRGYKREVPVLTQDFLIVISGAOECT
 KTRFERERQECYKAYLAIRQHANLFINLFSMILGSGMPELQSFDDIAYIRKTLALDK
 TEQEALEFEMQMDAHHGQWTKMDWIFHTIKOHALN"

BASE COUNT

1028 a 581 c 680 g 918 t

ORIGIN

Query Match 49.1%; Score 26; DB 3; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AGCATTTCTAAGTCACTATCATCC 35

Db 110 AGCATTTCTAAGTCACTATCATCC 85

RESULT 3

A37232/c

LOCUS A37232 3498 bp DNA PAT 05-MAR-1997
 DEFINITION Sequence 1 from Patent WO9403609.
 ACCESSION A37232
 NID 92294345
 VERSION A37232.1 GI:2294345

KEYWORDS unidentified.
 SOURCE unidentified.
 ORGANISM unidentified.

REFERENCE 1 (bases 1 to 3498)
 Parker, P.J., Goode, N.T., Nurse, P.M. and Waterfield, M.D.
 EUKARYOTIC CELLS TRANSFORMED WITH A MAMMALIAN PHOSPHOLIPID KINASE
 OR PROTEIN KINASE AND ASSAYS USING THEM
 Patent: WO 9403609-A 1 17-FEB-1994;
 IMP CANCER RES TECH (GB)
 COMMENT Other publication JP 8503124T 960409.
 FEATURES
 source location/Qualifiers
 1..3498
 /organism="unidentified"
 /db_xref="taxon:32644"
 1..3207
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAA02284.1"
 /db_xref="PID:e304808"
 /db_xref="PID:g2294346"
 /db_xref="GI:2294346"

CDS

translation="MPRRSGELMGHILMPRIIVLECLPNGMIVTLECLREATLIT
 IKHELFKARKYPLHQLODESSYIFVSTQAEAREEFDETRRLCDLRFQPLKVI
 EPVGRREKILNREIGFAIGMPCEDKPEVDPRNLIINCKEAVDLRLDSPH
 SRAMVYPPNVSSELPKHIYKRLDKGQIIYIVIVYSPNNOKKXTLKHHCQVPE
 OYIAEAIKRTSRMLSSBOLKCYLEVOGYIILKVCQDETFLEKPIPSQYKTRSC
 IMLGRPNMLMAKESLSQLPMDCTMPSTSRISTATPYNGETSTKSLMVINSAL
 RIKICATYVNVNIRIDIKIYRTGIIYHGGELCDNVNTQRPVCSNPRNMLNVDIY
 IPDLPRARLCLISCSYGRGAKKEHCPKANGINLFDYDTLVSGKMLNLPVPH
 GLEDLNPIGVTSNPKETPCLEPEPFMSYVYKPPMSYIEEHANVSREGEFSY
 SHAGISNRLARDNELRENDKQLRAICTRDLSETTDEKDFLSHRHVCYTIPEIIP
 KILLSYKNSDEVAQMTCLYKMDPIRPEBAMELDCNTPDPWVRGFAVCELYLT
 DDKLSQYLIQVQVAKTYEQYLDNLVRLKALNORIGHEFFHMLKSEHNNTVSO
 RFGLLSEYCRACGMYLKHLNROYAMEKLINLTDILKOEKDETOKVMKFLVQMR
 RPFMDALOGFSLNPAHQNLNLEECRIMSARPLMLNMPNDIMSELIFQNNH
 IIFKNGDRLQDMFLTIQIRIMENIMONOGDLMLPYGCLSIDCVGLIEVRSHT
 IMOICGKGLGALQPNSTLHOMLKXNKGEIYDAADLPTRSCAGCVATFTLLGIG
 DRHNSNIVKDDGOLFHDGFHFLDHKKKRGYKREVPVLTQDFLIVISGAOECT
 KTRFERERQECYKAYLAIRQHANLFINLFSMILGSGMPELQSFDDIAYIRKTLALDK
 TEQEALEFEMQMDAHHGQWTKMDWIFHTIKOHALN"

BASE COUNT

1147 a 626 c 722 g 1003 t

ORIGIN

Query Match 49.1%; Score 26; DB 5; Length 3498;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 AGCATTTCTAAGTCACTATCATCC 35
 ||||||||||||||||||||||||||||||||||||
 Db 110 AGCATTTCTAAGTCACTATCATCC 85

RESULT 4

HSPH13K/c

LOCUS HSPH13K 3424 bp mRNA PRI 24-AUG-1995
 DEFINITION H.sapiens mRNA for phosphatidylinositol 3-kinase.
 ACCESSION Z29090
 NID 9472990
 VERSION Z29090.1 GI:472990

KEYWORDS phosphatidylinositol 3-kinase.

AUTHORS Kllppel, A., Escobedo, J.A., Hirano, M. and Williams, L.T.
TITLE The interaction of small domains between the subunits of phosphatidylinositol 3-kinase determines enzyme activity
JOURNAL Mol. Cell. Biol. 14, 2675-2685 (1994)
REFERENCE 94187738
AUTHORS 2 (bases 1 to 3207)
TITLE Kllppel-Giese, A.
JOURNAL Direct Submission
 Submitted (05-NOV-1993) Anke Kllppel-Giese, CVRI and Dalilhi Research Center, University of California San Francisco, 505 Parnassus Ave, Long Hospital, San Francisco, CA 94143, USA
FEATURES Location/Qualifiers
 1..3207
 /organism="Mus musculus"
 /strain="Balb/c"
 /db_xref="taxon:10090"
 1..3207
 /standard_name="p110 protein"
 /codon_start=1
 /product="phosphatidylinositol 3-kinase 110 kDa subunit"
 /protein_id="AA18334.1"
 /db_xref="PID:941495"
 /db_xref="GI:41495"
 /translation="MPPPSGSELMGCIHMPRIIVCELLPNGMYTLECLREATLV
 IKHELFREARKYPLHQLQDETSYIFSVYTOEAREEPDEFDRICDLRLPQPLKVI
 EFGNREKILNREIGFVGMPCEPDWAKDEQDFRNINLVNCKEANDLDINSFH
 SRANYVPPNVSSPELPKHTINKDKGOIIVVIVTVSPNDKQITIKINHCVE
 QVIAEIRKTRSMILSSQLKCYLEVQKILVKGDEDFLEKRYLSQYKVRSC
 IMGRPMIMLMAKESLYSQLIDSTFMSYSRSRISTAPYKNGSTSLVYINSAL
 RKIKICATVYVNIKIDIKIVYRTGIGHGEGLCDVNPAPCSNPRENEMLVNIDY
 IPDLPRILALCISGVKGRKAKHEHCLANGNINLEPYTITLISGKALMLMPVH
 GEDLNLPGVGSNPNKPTCLELFPDSSVYVPEKMSVIEHANSVSEAGFVS
 SHTGSNRLARPNELRENDKEDRALCTDPSEITBOKRDLVSRHRCVITPELP
 KILLSKMSRDEVAQMTCLVDPWPIKPEQAMELDCVPPDFVMSFARVLELYLT
 DKLQYLLIOLVQVLYKEQYLDNLVRFILKALNQRIGHFFHMLKEMKNKYVQ
 RGLLSEYCGACGMYLKLNRQVEAMEKLINTDLKDEKDEQVOKMYELVQMR
 QDFMDALGGFLSPMLPAHQGLNLRLEGRINSKARPEMLMENDDIWSELFPQMR
 IIFKNGDDLRDMITLQIRIMENIMONGDLPMIPYCSLIGCVGLIVRVSHT
 IMQIOCKGLKKAIDPNSHTLQWMLKDKRKGELYDAIDLFTRSCGCVVAFITIG
 DHNSNINWKGDLPHIDEGHFLDKRKRKFRKRVFVLVQDFLIVISGAGEYTG
 TREFEREQEMCYKAYLAIRQANLEFINLSMGLSGMPELQSPFDIAIYIRKTLADK
 TEOEALEYETTKOMDANHGWTTKMDMIFHTIKOHLN"

BASE COUNT 958 a 693 c 734 g 822 t
UNIGIN

Query Match 46.0%; Score 24.4; DB 12; Length 3207;
Best Local Similarity 96.2%; Pred. No. 5.5;
Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 10 AGGCATCTAAAGTCACTATCATCC 35
 |||||||||||||||||||||
Db 110 AGGCATCTAAAGTCACTATCATCC 85

RESULT 7
MATGENB/C MATGENB 166924 bp DNA circular PLN 13-NOV-1998
LOCUS A.thaliana mitochondrial genome, part B.
DEFINITION Y08502
ACCESSION Y08502
NID 91785729
VERSION Y08502.1 GI:1785729
KEYWORDS 18S ribosomal RNA; 26S ribosomal RNA; 5S ribosomal RNA; apocytochrome b; atp1 gene; atp6 gene; atp9 gene; atpase subunit 1; atpase subunit 6; atpase subunit 9; cb203 gene; cb206 gene; cb256 gene; cb382 gene; cb452 gene; cob gene; cox1 gene; cox2 gene; cox3 gene; cytochrome c biogenesis orf203; cytochrome c biogenesis orf206; cytochrome c biogenesis orf256; cytochrome c biogenesis orf382; cytochrome c biogenesis orf452; cytochrome c oxidase subunit 1; cytochrome c oxidase subunit 2; cytochrome c oxidase subunit 3; mat1 gene; maturase; nad1 gene; nad2 gene; nad3 gene; nad4 gene; nad4l gene; nad5 gene; nad6 gene; nad7 gene; nad9 gene; NADH dehydrogenase subunit 1; NADH dehydrogenase subunit 2;

SOURCE ORGANISM
 Mitochondrion Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; Rosidae; Caprales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 166924)
AUTHORS Unsel, M., Marlenfeld, J.R., Brandt, P. and Brennicke, A.
TITLE The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366924 nucleotides
JOURNAL Nat. Genet. in press
AUTHORS 2 (bases 1 to 166924)
TITLE Marlenfeld, J.R.
JOURNAL Direct Submission
 Submitted (30-SEP-1996) J.R. Marlenfeld, Institut fuer Genbiologische Forschung GmbH, Innestrasse 63, 14195 Berlin, FRG
REFERENCE 3 (bases 1 to 166924)
AUTHORS Marlenfeld, J., Unsel, M., Brandt, P. and Brennicke, A.
TITLE Genomic recombination of the mitochondrial atp6 gene in Arabidopsis thaliana at the protein processing site creates two different presequences
JOURNAL DNA Res. 3 (5), 287-290 (1996)
COMMENT 97191539
MEDLINE Related sequences:
 X96535, X96536, X91201-X91210, X72616, X67736, X65123, X67105, X60045, X60047, X60048, M58320, X07376, X94583, X82565, U94902 & U94903.
FEATURES SOURCE
 1..166924
 /organism="Arabidopsis thaliana"
 /mitochondrion
 /variety="Columbia"
 /db_xref="taxon:3702"
 /clone_id="lorist X"
 /sub_clone="pnc19"
 /sub_clone="pnc18"
 /sub_clone="pbluescript SK"
 /sub_clone="pbluescript KS"
 1729..2097
 /gene="nad1"
 /note="orf122c"
 /codon_start=1
 /protein_id="CA65843.1"
 /db_xref="PID:e283572"
 /db_xref="PID:g1785730"
 /db_xref="GI:1785730"
 /db_xref="SPRMBL:p92510"
 /translation="MKNASFCLSRIILAPCRILGTLILLPYPCSTLRQFLFRLS
 LFRHVEWIPAGLSHHPYPLASPPVETLLIARLLLSIKQLSLPAPKASIASL
 DAKTRGRLSSCSYCPPT"
 complement(3588..4310)
 /gene="nad1"
 /note="orf140a"
 /codon_start=1
 /protein_id="CA69844.1"
 /db_xref="PID:e283452"
 /db_xref="PID:g1785731"
 /db_xref="GI:1785731"

/db_xref="SPTREMBL:P92511"
 /translation="MRSYVLRSLRGKRVINLESTRKRLRSTNIVPGRKKQKSIKSK
 NMAKKGPIILVLRGKRSDDSRSAELJCLYFFLYVAPTGLVLRLLHFW
 GLRLGLENHFHGVGPDGATGLDLPQEQDPTLGVNRALDLNELPVLHVE
 VEEPOSTKQANDVLAHLNQVQLTRDLQTEPNIMRQALDILDMEVRSLQRFRRLF
 RQRDLREYQVRSMLREQLNRYR"
 complement(4340..4412)
 /gene="trna-lys"
 /product="transfer RNA-lys"
 complement(4340..4412)
 /gene="trna-lys"
 complement(7553..7915)
 /gene="nad1"
 /note="orf120"
 /codon_start=1
 /protein_id="CAA69816.1"
 /db_xref="PID:e283453"
 /db_xref="PID:g1785732"
 /db_xref="GI:1785733"
 /db_xref="SPTREMBL:P92512"
 /translation="MNRITIEKVSMLCEGCLPPTPRADAANTAVHIINRYSTAINF
 HVDVWFQSGVPTSYLRRFCVAYIHCDGSKLKPRAKGEKSYLINRIVSILYTI
 GIGKTSPPRAKSHLGING"
 9499..9822
 /gene="nad1"
 /note="orf107d"
 /codon_start=1
 /protein_id="CAA69817.1"
 /db_xref="PID:e283573"
 /db_xref="PID:g1785733"
 /db_xref="GI:1785733"
 /db_xref="SPTREMBL:P92513"
 /translation="MEDFGFTTALRSVYIVVERISSSLSYNTLSSAIFRLDWIKT
 PRHRAARALGIONSGICTACPHLSALFGLSVASCPSKNDGLHTRIFYDFEFP
 DSAVF"
 17631..17639
 /gene="cox3"
 /evidence="not_experimental"
 17631..19077
 /gene="cox3"
 18280..19077
 /gene="cox3"
 /codon_start=1
 /evidence="not_experimental"
 /product="cytochrome c oxidase subunit 3"
 /protein_id="CAA69818.1"
 /db_xref="PID:e283454"
 /db_xref="PID:g1785734"
 /db_xref="GI:1785734"
 /db_xref="SWISS-PROT:P92514"
 /translation="MISQSRSHYLVDPSPWPISGSLGALATTVGVMYMHPEGGAR
 LLSLGLIFLYTMEVWRDYLRESLLEGHHTKVVQLPRYSILFVSEVFFPAFPM
 ASSHSLAPAVEIGIWPKGIEVLDPWEIPFLNTPLPSGAIVTAHHHLLGKKEK
 RAYALVATVLAIVTFQGMEXYQAPFTISDSYGSTFLAGFPGFHHYIICLTLL
 ITCIRYLIHLTRENHVGFEAAMVHFVDVWVLFVSLIYMGGI"
 complement(20471..20773)
 /gene="nad1"
 /note="orf100a"
 /codon_start=1
 /protein_id="CAA69819.1"
 /db_xref="PID:e283574"
 /db_xref="PID:g1785735"
 /db_xref="GI:1785735"
 /db_xref="SPTREMBL:P92515"
 /translation="MLHISOEFLPLISMEHRLITSOVIKKIGIPKVIHFARONN
 RGLILOIFHHIDFLGFAIIVYVLDKSRQVINDSILNSLVLYNQKLLKEK"
 20830..21189
 /gene="nad1"
 /note="orf119"
 /codon_start=1
 /protein_id="CAA69820.1"
 /db_xref="PID:e283455"
 /db_xref="PID:g1785736"
 /db_xref="GI:1785736"

/db_xref="GI:1785736"
 /db_xref="SPTREMBL:P92516"
 /translation="MHRKSEGHSLPLCMHRSMODISCHLKOMPFEVLQAGFYWPT
 TFDKADHFGVSCACORKNFTKRNEMPOHFIIEVEFDWGIYFNKRTLSMKRPIHP
 NGRGLCKMGSSCEPH"
 complement(21700..22029)
 /gene="nad1"
 /note="orf109b"
 /codon_start=1
 /protein_id="CAA69821.1"
 /db_xref="PID:e283456"
 /db_xref="PID:g1785737"
 /db_xref="GI:1785737"
 /db_xref="SPTREMBL:P92517"
 /translation="MLRSKLELEEVVDDKEEDQVYVRKAKQIVNPEPLGLMSARIG
 LRWPFVTKCSPLTSOSTAPYGTSPMRKRLDLNPGNNLSLELRSTYSFTLN
 SLGRIFL"
 complement(22687..22989)
 /gene="nad1"
 /note="orf100b"
 /codon_start=1
 /protein_id="CAA69822.1"
 /db_xref="PID:e283575"
 /db_xref="PID:g1785738"
 /db_xref="GI:1785738"
 /db_xref="SPTREMBL:P92518"
 /translation="MGLSTGCOLVFSKRPMTNTHVSPAMEGSLKYPFPNINITSKV
 SIKLNDRYNLLYDEGRVRLSLNLNGERCDEVPELSEVLEFPLQSSGYEKLEHSNN"
 27087..27619
 /gene="nad1"
 /note="repeat 11"
 27109..27196
 /gene="trna-ser"
 /product="transfer RNA-Ser"
 27109..96784
 /gene="trna-ser"
 27502..27577
 /gene="trna-tyr"

Query Match 45.38; Score 24; DB 7; Length 166924;
 Best Local Similarity 75.08; Pred. No. 7.2;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 10 AGGCATTCTAAGTCATCATCCCGACTCGAGTCGACA 49
 DB 58270 AGGTATCTCTAAATCAAGATATCCCGACGTGAGCAGTCA 58231

RESULT 8
 SCYOR007C
 LOCUS
 DEFINITION
 ACCESSION
 NID
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 baker's yeast.
 Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
 Saccharomycetaceae; Saccharomyces.
 1 (bases 1 to 2656)
 Pettersson,B., Sterky,F. and Uhlen,M.
 Unpublished
 REFERENCE
 2 (bases 1 to 2656)
 MIPS.
 AUTHORS
 TITLE
 JOURNAL
 Direct Submission
 Submitted (04-JUL-1996) Data collected by MIPS on behalf of the
 European yeast chromosome XV sequencing project. MIPS at the
 Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
 FEATURES
 source
 1..2656
 /organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"
 /chromosome="XV"
 complement(317..1357)
 /note="ORF YOR007c"
 /codon_start=1
 /protein_id="CAI99195.1"
 /db_xref="PID:g251943"
 /db_xref="PID:g1420100"
 /db_xref="GI:1420100"
 /db_xref="SPTREMBL:Q12118"
 /translation="MSASKEELIALIVNFSSIVEKEISEGADSLNVAACDISFAF
 GEREAVSGLGSEFGQGLADILNSASRVPSNKKDAENVEINIPEDDETAKA
 EDLKGKNGKAMANKDELAINTYTAIRYLAAYARAASHSIKRYDDAYDAE
 SAISIDPSYFRGYSRIGPAKYAGKREALEAKYVLDIEGDNATAMKRDYESAKK
 VQSLNLEKTVPEQSDADVDAQSAGSLPDVGLGGLGGLGGLMNNPQLMQAAQK
 MSNPGAMQNIQKMDPSIRMAEGFASGGTPTNLSIDMNNPALRNMAQNLFGAGAQ
 STDEPNDENKK"
 complement(1678..1750)
 /gene="tr(ACG)OR - systematic name"
 /note="anticodon gene: ACG; tRNA-Arg2 - common name"
 complement(1678..1750)
 /gene="tr(ACG)OR - systematic name"
 complement(1949..2273)
 /note="delta"
 BASE COUNT 819 a 491 c 426 g 920 t
 ORIGIN

Query Match 45.3%; Score 24; DB 7; Length 2656;
 Best Local Similarity 68.8%; Pred. No. 8;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 4 AACCTAGGATCTTAAGTCACTATCATCCGACCTGAGTCGACATC 51
 DB 602 AAGATCTGGCAATCCACCCGCTAGCACCCTGATGATCGACATC 649

RESULT 9
 LOCUS ATAC007143 75812 bp DNA PLN 05-JUN-1999
 DEFINITION Arabidopsis thaliana chromosome II BAC T17H1 genomic sequence,
 complete sequence.
 ACCESSION AC007143
 NID 95001456
 VERSION AC007143.3 GI:5001456
 KEYWORDS HTG.
 SOURCE Thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 Arabidopsids.
 REFERENCE 1 (bases 1 to 75812)
 AUTHORS Lin,X., Kaul,S., Shea,T.P., Fujii,C.Y., Shen,M., Vanaken,S.E.,
 Barnstead,M.E., Mason,T.M., Bowman,C.L., Ronning,C.M., Benito,M.,
 Carreira,A.J., Creasy,T.H., Buell,C.R., Town,C.D., Nierman,W.C.,
 Fraser,C.M. and Venter,J.C.
 Arabidopsis thaliana chromosome II BAC T17H1 genomic sequence
 Unpublished
 TITLE 2 (bases 1 to 75812)
 AUTHORS Lin,X. and Kaul,S.
 JOURNAL Direct Submission
 Submitted (23-MAR-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA, xlin@tigr.org
 Lin,X.
 REFERENCE Direct Submission
 Submitted (05-JUN-1999) The Institute for Genomic Research, 9712
 Medical Center Dr., Rockville, MD 20850, USA
 On Jun 5, 1999 this sequence version replaced gi:4512639.
 Address all correspondence to:
 Xiaoying Lin
 The Institute for Genomic Research

9712 Medical Center Dr.
 Rockville, MD 20850, USA
 e-mail: xlin@tigr.org
 BAC clone T17H1 is from Arabidopsis chromosome II and is near the
 molecular marker m1421.
 The orientation of the sequence is from SP6 to T7 end of the BAC
 clone.

Genes were identified by a combination of three methods: Gene
 prediction programs including GRAIL (available by anonymous ftp
 from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
 Washington), GenScan (Chris Burge,
 http://genomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
 (http://www.cds.dtu.dk/netgene/cbsnetgene.html), searches of the
 complete sequence against a peptide database and the Arabidopsis
 EST database at TIGR (http://www.tigr.org/tdb/at.html).
 Annotated genes are named to indicate the level of evidence for
 their annotation. Genes with similarity to other proteins are named
 after the database hits. Genes without significant peptide
 similarity but with EST similarity are named as 'unknown proteins'.
 Genes without protein or EST similarity, that are predicted by more
 than two gene prediction programs over most of their length are
 annotated as 'hypothetical' proteins. Genes encoding tRNAs are
 predicted by tRNAscan-SE (Sean Eddy,
 http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats are
 identified by repeatmasker (Arian Smit,
 http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of
 genomic sequence that are not annotated as genes but have predicted
 exons by GRAIL are annotated as misc features.

The entire sequence of T17H1 is nearly identical to the published
 mitochondrial genome. Since we cannot be certain that the genes
 found in this BAC are expressed from the nucleus, we have chosen
 not to annotate this BAC.

FEATURES

source
 Location/Qualifiers
 1..75812
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /map="m1421"
 /chromosome="II"
 /clone="T17H1"
 complement(1..5719)
 /note="Overlap with BAC clone T5M2 (1..5718)."
 /note="Overlap with BAC clone T5M2 (1..51765)."
 complement(1..56697)
 /note="matching Arabidopsis ectotype C24 mitochondrial
 genome, part B (GB:Y08502:17688..74459)."
 complement(58488..66262)
 /note="matching Arabidopsis ectotype C24 mitochondrial
 genome, part B (GB:Y08502:9924..17696)."
 complement(66259..75812)
 /note="matching Arabidopsis ectotype C24 mitochondrial
 genome, part B (GB:Y08502:134052..143607)."
 misc_feature
 misc_feature
 misc_feature
 BASE COUNT 20824 a 17017 c 16566 g 21405 t
 ORIGIN

Query Match 45.3%; Score 24; DB 8; Length 75812;
 Best Local Similarity 75.0%; Pred. No. 7.3;
 Matches 30; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 QY 10 AAGCATTTAAAGTCACTATCATCCGACCTGAGTCGACA 49
 DB 16203 AAGTATTTAAATCAAGATATCCGACGAGCA 16242
 RESULT 10
 LOCUS SCU43491 30469 bp DNA PLN 25-OCT-1996
 DEFINITION Saccharomyces cerevisiae cosmid clone pEOA156 from chromosome XV.
 ACCESSION U43491 X95109
 NID 91150992

VERSION U03491.1 GI:1150992
 KEYWORDS Baker's yeast.
 SOURCE Saccharomyces cerevisiae
 ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 REFERENCE 1 (bases 1 to 30469)
 AUTHORS Sterky,F., Holmberg,A., Pettersson,B. and Uhlen,M.
 TITLE The sequence of a 30 kb fragment on the left arm of chromosome XV from Saccharomyces cerevisiae reveals 15 open reading frames, five of which correspond to previously identified genes
 JOURNAL Yeast 12 (108 suppl), 1091-1095 (1996)
 MEDLINE 97051599
 REFERENCE 2 (bases 1 to 30469)
 AUTHORS Sterky,F. and Uhlen,M.
 TITLE Direct Submission
 JOURNAL Submitted (14-DEC-1995) Fredrick Sterky, Department of Biochemistry and Biotechnology, Royal Institute of Technology, Teknikringen 34, Stockholm, 100 44, Sweden
 COMMENT On Apr 10, 1996 this sequence version replaced gi:1237089.
 FEATURES
 source
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /chromosome="XV"
 /clone="cosmid clone pEOA156"
 /gene="UNE378"
 complement(127..1263)
 /gene="UNE378"
 complement(127..1263)
 /note="ORF UNE378, unidentified function"
 /codon_start=1
 /product="hypothetical protein UNE378"
 /protein_id="AAC49477.1"
 /db_xref="PID:g1150993"
 /db_xref="GI:1150993"
 /translation="MPYKLRPMVIGIALPFLISFICYGAGHAFILSNFLSPKQITE
 FCLSMITSLYLAICTNPGRLPYTKPPDIWRFCKCOSYKPERSHHCTCQCVL
 MMDHCHWIMNVCVFANYPHFLRFLEWIVTTSVLCIQARIYFIQQRHLPQYFK
 KSELIPLTSSPLSEFVLLITLILFRLCLFNQILNGSISMDRLSESLFNGRLT
 OKLIDNTWRIAYEERSFQNKKADEHLTKRPREDELVNPEDYDELTYNALLYGPPIH
 LMIMPYGVPTGDSNPFKNGISKYEANSLSLEDTLSLPMPDGGKTNTVFNHGSSTIE
 MRNMSGQLIRTPRPONGRNASRKTWNDSGLDDGVVDME"
 complement(1680..2663)
 /gene="UND327"
 complement(1680..2663)
 /gene="UND327"
 /note="ORF UND327, unidentified function"
 /codon_start=1
 /product="hypothetical protein UND327"
 /protein_id="AAC49478.1"
 /db_xref="PID:g1150994"
 /db_xref="GI:1150994"
 /translation="MLHVGESFOMSTLLERTKSVOLIKKRAAKTSANPAVAKAK
 VLRLYSMDPEWQRNDPRLIHGYETSFIETPSLFLHNSVNTSHLPAAG
 FFLVLLDKSTIKFAFTIWDHNVIDLFTSGAFACILSSFFQLSHSLRATLEN
 KLYLIGICILIVTSVAILLYGYEKLSEFLALIVSFACISIVLSKDKFERRE
 RPYAGLFFVCGLSIIPISGLCYFSEFWIOQLFVLLGGVLIIGVILGAMRF
 PERKPGKFDIMGSHQLHFLLVIVIALCHLRLGLNSYELVHIMENGIVS"
 complement(3518..4339)
 /gene="PHO80"
 /db_xref="SGD:S0005361"
 3518..4339
 /gene="PHO80"
 /note="Pho80"
 /codon_start=1
 /db_xref="SGD:S0005361"
 /product="phosphate system cyclin"
 /protein_id="AAC49479.1"
 /db_xref="PID:g1150995"
 /db_xref="GI:1150995"
 /translation="MESTSGRSENHEDGIPKVIIPADFNKSKRDIIVYLISRMVY
 SLAIINENSATKKSDQITLIRYHSKIPNISINIVIRLTKSSLECHVAMTSLYIT

DILOTVYPDTFINSLTARHRLTATTATATGCLDSESTNAHYAKGVCHENLTLEN
 DFLKRVNRIPLRDNNTLCSIEQKKKPVLDKNAISLDLDSTSYNRRPSGYNVD
 KYRRIIVQLVGSFNPDKSRKVDYVLPPIIDIVSEGSQTOLKSSSPNHSQKR
 YSEAKDAHIYNRKRPD"
 5101..7302
 /gene="UNC733"
 5101..7302
 /gene="UNC733"
 /note="ORF UNC733, unidentified function"
 /codon_start=1
 /product="hypothetical protein UNC733"
 /protein_id="AAC49480.1"
 /db_xref="PID:g1150996"
 /db_xref="GI:1150996"
 /translation="MTSENPVLLSRVINYVRAASSIASODVDFYKNMLDRGFSKDLK
 KADRLAMANEIILSIDENHSEFLKEDSIDLWNNGNIMDNLNLSMDLKNKA
 INSKRSGSDLOVGEFSGKNPSPKREKOLKRSQIDNSSEPFPLKERNMAK
 PLSLSLVDDDENPESHYHPTEYEDHDSYISYGLVCLMOISREYDLDTUKRLNLI
 STELESMLDKNTKEIYAVDLEHHDYRSGYGLVCLMOISREYDLDTUKRLNLI
 LNEFTNPSTVYKVFHGAFMQIMLQRLAGYVGLFDTYHASKAIGLRRSLAILE
 FANFTSKRYQLADMRIPLSKPMTAYARADTHELNIYDQLRKLIENKLAGVLE
 SRNVAKRREYSKRYPLTSESEVSPLEKSPKILMYOINIPEREVLVRELQMD
 LIARDESPREVPNOLLAIVAYPTDVIYGVSLNGTEHVRONAKLIANIRBA
 LRNTKNNBEATPSPSEETADGILLETISVPOIRDMERSVCNSNISKSRAPPT
 NSLTLLKILPREHDIATSRDGLPNKVKTEDIRAKONRSALANEDLIFETKPL
 VVPKLEIKITVDPASAPNHSPEIDNDDLVYLNKKNIQKQPAKREKGYEKDADVS
 KIPNLSNKPQGNRRQKKRRRFPSSSDSNCPRAKKRRRAKKNLSFR"
 7686..9320
 /gene="UNA544"
 /db_xref="SGD:S0005528"
 7686..9320
 /gene="UNA544"
 /note="ORF UNA544, unidentified function"
 /codon_start=1
 /product="hypothetical protein UNA544"
 /protein_id="AAC49481.1"
 /db_xref="PID:g1150997"
 /db_xref="GI:1150997"
 /translation="MAIGKRLLVNKPAAESFVSPMYDPLPPRVGNOMLPEYIIF
 CAVLRCTIGLGPYSGSPPLKEDGPEAOSHMMETIHOPLSKWYVLDIYWGIDVPP
 LTAHSLILIGISGFNPSPFALEKSGFSPNDGLTTRKSTVYISDILFPYAVY
 FTKWIGRYNOSPITGOSIASALLIPOSMLIDBGHVOYNSVGLTAVAINNLDDY
 YAAAVCFVSLICFKOMALYAPAFEFVILSLRSLEFKFPIARLTVAFTALFAII
 FAPLYFAGGLKINHOCIRIFRPFARIFEDKRVANFCVNVFVYKKEFTIOLQAI
 SLIATVIGFLPAMITLHPKHLPLVILACGSPFLSFQVYEKTLILPLDTLL
 YSSIDVNVILSVMINNVALLFTMLPKKGLKHQVAVFLSSMLKGNRSFTIPREL
 PKSLTPSPSISINSDDRSLDPTNYNVMKSFITIGYIANGFYHFLDQVAPPSKTPD
 LNVILNCAVGFICSFIFLWWSYKIFITSGSKMKDL"
 9724..11160
 /codon_start=1
 /product="subtilisin-like protease III precursor"
 /protein_id="AAC49482.1"
 /db_xref="PID:g1150998"
 /db_xref="GI:1150998"
 /translation="MKFSTLPIIMANCCMIIPDEGIVRFIENIDGTRSVAGSG
 LGQDHPNFTTEHGVNAHKEFLPYRVYIENEDISQQLQSHQVVOXHSVSGVL
 TENDAFNRVSSSVSSKSGOPGIDNPFIDNGLEFGYGFETDIKIIISODPIIKVE
 OETVAKISNSISLEAPWGLARVHNSRPRKYGODLFEVLDAGKGVTSVLDGIDT
 EHDPEGRAGAVIIPANDASDLNGHTCAIGITGSKHGVAKNTIVAVKYLRSNG
 EGVIVSDVIKGIETVTKHEISSKRNKEFGSTANLSLGSSKSLAMEMAAVAAVDSGV
 HFLAIAGNEDEDACLSPPAGAEKSIYGAFTESDDRAFFSNMGCVDFVAFGINIMST
 YIGSRNATLSLGSFASAPVAGILSLQAPADSEFPNDAPSPQELKEKYLKFS
 OGVLGDIDDPNKLINGGKKLIDGW"
 11161..12625
 /gene="UNB254"
 11161..12625
 /gene="UNB254"
 /note="ORF UNB254, unidentified function"
 /codon_start=1
 /product="hypothetical protein UNB254"
 /protein_id="AAC49483.1"
 /db_xref="PID:g1150999"

```

/db_xref="GI:1150939"
/translation="MKRAKSYRKQLVYSHTFREPIOVLDNQLVLECNSEN
LPSGLKRLQADVAVITQCCIALYERNDGAILAOFERRCNHSEFKDPKSAEC
IESVNIIGANKHRYVVAQODIDLRKRLTVPGVYLIILTSVMYMLSTASASK
ITEBQKLYGLNDPNIEKLOESDGSSESIRKRLKLPKAPNPLSVKKKKVNPSPD
EYKEDETSKEKKRRRRKHSNTNPVSNGTAAQ"
complement(12778..14136)
/gene="UNE452"
complement(12778..14136)
/gene="UNE452"
/note="ORF UNE452, unidentified function"
/codon_start=1
/product="hypothetical protein UNE452"
/protein_id="AAC49484.1"
/db_xref="PID:9151000"
/db_xref="GI:1151000"
/translation="MGLGLVLDDEEYKRRHOGDSEIVDHSOEKHIONSRKRLKS
FCSIANGISOEFEKIDRKTGHMKRTEVAPPAPILIEGSKIPAEWIDPSSTVLEI
KSRSDNETNMOKATNCTLYGCKRIKRYDKETDCTYLDLTSRVSNSPYOA
ERSDGLIRKRRKRLVLSDFHONKQPIPSIRPGLFYVLSDYEDTGIRITRAE
LEKTIYEGKLLIVILKHSIGDVRILISCTTTECKALIDRGDILHAPMWLDICIA
YKRLILEPNYCFNVSQMRAVAKRVDCLDSFENDISEKLSLSYKSQLSLPMGE
LEIDSEVRFPLFLFSNRIAYPRKISTEDDIEMKIKLFGKITTDOOSCNLIIP
YTDPLRKDCMNEVEHEKIKEQIKASDTIPKIAVAVAPMVDSINENQVPEDEPPV
NY"
complement(14387..15610)
/gene="UND407"
complement(14387..15610)
/gene="UND407"
/note="ORF UND407, unidentified function"
/codon_start=1
/product="hypothetical protein UND407"
/protein_id="AAC49485.1"
/db_xref="PID:9151001"
/db_xref="GI:1151001"
/translation="MISALDSIPPEQNPAPSPDFKWLCEELFKIHEYOINGTAGTK
SRSEFYIISNFEMMRKTVGNNTYPAVLALPRDRIRYNIDYVILRICSVILK
PKNSATEORLMDKORVCGNLSLVEIARAPESKAITIDNNHILDSISGD
RPSGRGKSLVSKSPFLACVENSVELKTFEDIVLNRYIGOEHLNLCWHPDAQ
DYLVSISDLKAVTSKLDPKVRLKDDSLVIGFAFAPQAKVNLSEKICRLHLD
FLVEKMDGERIOVHYNMGESIKFFSRGIDYTLXGASLSGTSISQHLFTSVKE
CULDGEWTFDAKRVRILPFGLVKSAKALSFNINNVDFHLYMVDLLXNGTSL
TPLPHORKOYLNSI"
complement(15947..16888)
/gene="UND313"
complement(15947..16888)
/gene="UND313"
/note="ORF UND313, unidentified function"
/codon_start=1
/product="hypothetical protein UND313"

Query Match 45.3%; Score 24; DB 8; Length 30469;
Best Local Similarity 68.8%; Pred. No. 7.5;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 4 AAGCTTAGGATCTTAAGTCACTATCATCCGACGATCGATGCATC 51
Db 17490 AAGATCTGGCATCCACCGCGACCTAGCACCTGACTGCATGCATC 17337

```

```

AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
Boore,J.L. and Brown,W.M.
Complete DNA sequence of the mitochondrial genome of the black
chiton, Katharina tunicata
Genetics 138 (2), 423-443 (1994)
95129806
2 (bases 1 to 15532)
Boore,J.L.
Direct Submission
Submitted (17-May-1994) Jeffrey L. Boore, Cell Biology and
Neuroanatomy, University of Minnesota, 321 Church St. SE,
Minneapolis, MN 55455, USA
FEATURES
source
Location/Qualifiers
1..15532
/organism="Katharina tunicata"
/mitochondrion
/db_xref="taxon:34587"
1..1548
/note="followed by 'T' as an abbreviated stop codon"
/codon_start=1
/translation="MGLGLVLDDEEYKRRHOGDSEIVDHSOEKHIONSRKRLKS
FCSIANGISOEFEKIDRKTGHMKRTEVAPPAPILIEGSKIPAEWIDPSSTVLEI
KSRSDNETNMOKATNCTLYGCKRIKRYDKETDCTYLDLTSRVSNSPYOA
ERSDGLIRKRRKRLVLSDFHONKQPIPSIRPGLFYVLSDYEDTGIRITRAE
LEKTIYEGKLLIVILKHSIGDVRILISCTTTECKALIDRGDILHAPMWLDICIA
YKRLILEPNYCFNVSQMRAVAKRVDCLDSFENDISEKLSLSYKSQLSLPMGE
LEIDSEVRFPLFLFSNRIAYPRKISTEDDIEMKIKLFGKITTDOOSCNLIIP
YTDPLRKDCMNEVEHEKIKEQIKASDTIPKIAVAVAPMVDSINENQVPEDEPPV
NY"
/db_xref="GI:1150939"
/db_xref="PID:9151000"
/db_xref="GI:1151000"
/translation="MKRAKSYRKQLVYSHTFREPIOVLDNQLVLECNSEN
LPSGLKRLQADVAVITQCCIALYERNDGAILAOFERRCNHSEFKDPKSAEC
IESVNIIGANKHRYVVAQODIDLRKRLTVPGVYLIILTSVMYMLSTASASK
ITEBQKLYGLNDPNIEKLOESDGSSESIRKRLKLPKAPNPLSVKKKKVNPSPD
EYKEDETSKEKKRRRRKHSNTNPVSNGTAAQ"
complement(12778..14136)
/gene="UNE452"
complement(12778..14136)
/gene="UNE452"
/note="ORF UNE452, unidentified function"
/codon_start=1
/product="hypothetical protein UNE452"
/protein_id="AAC49484.1"
/db_xref="PID:9151000"
/db_xref="GI:1151000"
/translation="MGLGLVLDDEEYKRRHOGDSEIVDHSOEKHIONSRKRLKS
FCSIANGISOEFEKIDRKTGHMKRTEVAPPAPILIEGSKIPAEWIDPSSTVLEI
KSRSDNETNMOKATNCTLYGCKRIKRYDKETDCTYLDLTSRVSNSPYOA
ERSDGLIRKRRKRLVLSDFHONKQPIPSIRPGLFYVLSDYEDTGIRITRAE
LEKTIYEGKLLIVILKHSIGDVRILISCTTTECKALIDRGDILHAPMWLDICIA
YKRLILEPNYCFNVSQMRAVAKRVDCLDSFENDISEKLSLSYKSQLSLPMGE
LEIDSEVRFPLFLFSNRIAYPRKISTEDDIEMKIKLFGKITTDOOSCNLIIP
YTDPLRKDCMNEVEHEKIKEQIKASDTIPKIAVAVAPMVDSINENQVPEDEPPV
NY"
complement(14387..15610)
/gene="UND407"
complement(14387..15610)
/gene="UND407"
/note="ORF UND407, unidentified function"
/codon_start=1
/product="hypothetical protein UND407"
/protein_id="AAC49485.1"
/db_xref="PID:9151001"
/db_xref="GI:1151001"
/translation="MISALDSIPPEQNPAPSPDFKWLCEELFKIHEYOINGTAGTK
SRSEFYIISNFEMMRKTVGNNTYPAVLALPRDRIRYNIDYVILRICSVILK
PKNSATEORLMDKORVCGNLSLVEIARAPESKAITIDNNHILDSISGD
RPSGRGKSLVSKSPFLACVENSVELKTFEDIVLNRYIGOEHLNLCWHPDAQ
DYLVSISDLKAVTSKLDPKVRLKDDSLVIGFAFAPQAKVNLSEKICRLHLD
FLVEKMDGERIOVHYNMGESIKFFSRGIDYTLXGASLSGTSISQHLFTSVKE
CULDGEWTFDAKRVRILPFGLVKSAKALSFNINNVDFHLYMVDLLXNGTSL
TPLPHORKOYLNSI"
complement(15947..16888)
/gene="UND313"
complement(15947..16888)
/gene="UND313"
/note="ORF UND313, unidentified function"
/codon_start=1
/product="hypothetical protein UND313"

CDS
trna
CDS

```

tRNA
LIVSLYSDHPY"
complement(3758..3822)
/note="codon recognized: TTC"
/product="tRNA-Phe"
/anticodon=(pos:3789..3791,aa:Phe)
complement(3821..3536)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 5"
/protein_id="AAC48366.1"
/db_xref="PID:9557278"
/db_xref="GI:557278"
/translation="MKAMFLISQATVLLISLSLSESTAMTTLITNKILIEFCPTN
ENSVITLPIILMTALGSAIVCYISAVMTETKYNMSODIEIRFTIYVWLFLVLSM
NLILFIPSLATLLQMDGLGVSCFLVYVYONKYSQAGLMTAFNNRIGDAVLAIS
VTFSGHFNALFEMSTPTMTLISMLLAGMKRSQALPFCSSWLPAMAAPVPSALVH
SSITVAGVFLIIQFPPLNSGHLSTSLMITSMTWAGISANFEDIKKILALST
LSQGVMLLSVSGILFSLALFLLPHAMKALFICAGNIISHNNODIRKSHMT
OMPSTCFNTANLALCGPPFAGFYISKDVIIEMAFNONNFLISLNFATILASY
SARSLFLFWSKMQSSVSTSDENSIIPATMLAAGATSGAVFSWLILPTSLPL
ISMNKLMAIFATLGAATYKLFLEKISHSIINFEFNMMFMINLSNNMTSKSSL
YGLYTFKSLDGMETIGEGILNKIMTLITINOKNOSNLEFLISLISMLFISV
TMWET"

CDS
complement(5537..5603)
/note="codon recognized: CAC"
/product="tRNA-His"
/anticodon=(pos:5568..5570,aa:His)
complement(5597..6928)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_except=(pos:complement(6788..6790),aa:Trp)
/transl_table=5
/product="NADH dehydrogenase subunit 4"
/protein_id="AAC48367.1"
/db_xref="PID:9335324"
/db_xref="GI:3335324"
/translation="MSRSTVTKNSKMKMYKMSVISHMFMTLNFSPNIRKIN
TEWSDNFSVPLMLSCWITGLIMSSYKILKNNLVSFLLNVLNFIIMVTK
SLFLYIFFEASLIPILLILIMWGYPQLQGMVMTYTLIGALPFLINFIETSHN
AHLNLLMMSLPIIMPYQAMISFWMLFILLVFKPIYSFHLMPKRAHEAVAGSMI
LAALLKLGYGINTINNINIKRYNFVNLFFITLGVGYSLLICIQSDMKALIAV
SSVGMGMLAGLISSPREGKMLIMIAHLSGSGFSSNMNMYEKSGSGLITK
CPISLVPTFSMSFELMCSTINMAAPSINILSDAGLISVYSKFLIMLALMAVSA
VYTLFTNTDGHKISSFVNPNPNKPLYNYVIFLHWIPAOQLILLSTTIC"

CDS
complement(6940..7242)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 4L"
/protein_id="AAC48368.1"
/db_xref="PID:9557280"
/db_xref="GI:557280"
/translation="MMNFENPLFTLGALSISALMTICLRKHLNTLLALEILWVSI
FLIYLSINIMSEGLMFVLLTLAASEISGLAMVLILIRKGDYVSSISIHNC"
7247..7312

tRNA
/note="codon recognized: ACA"
/product="tRNA-Thr"
/anticodon=(pos:7275..7277,aa:Thr)
complement(7312..7378)
/note="codon recognized: TCA"
/product="tRNA-Ser"
/anticodon=(pos:7348..7350,aa:Ser)
complement(7381..8520)
/codon_start=1
/transl_except=(pos:complement(8176..8178),aa:Trp)
/transl_table=5
/product="cytochrome b"
/protein_id="AAC48371.1"
/db_xref="PID:9335325"
/db_xref="GI:3335325"
/translation="MLKPIRSSHPVKIINGSLIDLPSPNSLIMWFGSLGLCII
OILGFLAMHYATNIELAFSSVAHICBRYNAGWLJRAHAGGSLFPCYFHHGRG
MYGGSFVNHTMWGVILLFLMGTAFIGIYVPLQCMGSMGATVITNLSAIPYIGTT

CDS
MYQWMSGFANDNALTREFSFHYLIAPPIIAASVYHILFIHETGNNPLGINSSEK
VTHSEFTLKDVGVFVNAVITLLVLFDLPDLAPDFIPANPLVFTHQPEWTF
FAVALRSQIPKRLGVGVGLMSIVILEFVPLINAKKEFTISFVPTQILYMSLVFL
LITWIGAQVPEVDPEFLIIGLITFVYFLFLFIMIPILTKMMDWTINE"
complement(8513..9019)
/note="Followed by 'T' as an abbreviated stop codon"
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 6"
/protein_id="AAC48369.1"
/db_xref="PID:9557281"
/db_xref="GI:557281"
/translation="MTFLITMSLFSQAFSIPULTOPISLILFMSLFISILFRC
SSWFSFLFVLIIGGLVMPFAYALPMLFKNLIQVFPVSTVWMLMLTUSD
IGTLIEQTKMFLMNSFNNHGISLFSFPLLITIVSLALIFVLLICVAKICFSNG
PLRPYKYA"

tRNA
9024..9089
/note="codon recognized: CCA"
/product="tRNA-Pro"
/anticodon=(pos:9054..9056,aa:Pro)
complement(9082..10032)
/codon_start=1
/transl_table=5
/product="NADH dehydrogenase subunit 1"
/protein_id="AAC48370.1"
/db_xref="PID:9557282"
/db_xref="GI:557282"
/translation="MFLFSMMVVIAYISILAVAFETLERKGLGYIQRRKPNKVS
IGLAPLADALKFLKENNLNYSNKLRYAPVLSLISCLMSLWSDPTLWYFN
GAPFLICSAISVYCTMTAGASNSKVALICAPVACTISFVYTLFVILSSAALS
SLNLOISSHOLITNYSILMTPLMLIYVCLAEITNRPAPBAGESLVSQFNIEG
SGSEFALFLAEVASILFMSMMTILFELGASWILTNLFWKTYFLAFVITWAGTF
PLRLYDLMLKLTWLSFLPSVSLIMMNLMSARFMTLLLS"

tRNA
complement(10033..10101)
/note="codon recognized: TTA"
/product="tRNA-Leu"
/anticodon=(pos:10068..10070,aa:Leu)
complement(10101..10165)
/note="codon recognized: CTA"
/product="tRNA-Leu"
/anticodon=(pos:10132..10134,aa:Leu)
complement(10166..11440)
/note="The exact endpoints are uncertain"

tRNA
Query Match 44.9%; Score 23.8; DB 37; Length 15532;
Best Local Similarity 72.1%; Pred. No. 9.2;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 1 TTTAGCTTAGCATTTCTAAGTCACTATCATCCGACCTCGAG 43
||||| ||| | ||||| ||| |||||
Db 10306 TTTAGCTTCTGCACTTAAAGTTACTTAAACCAACATCGAG 10348

RESULT 12

CHU68040 9829 bp DNA PLN 16-SEP-1996
LOCUS
DEFINITION
Cochliobolus heterostrophus polyketide synthase (PKS1) gene,
complete cds.
ACCESSION
U68040
VERSION
91546071
KEYWORDS
U68040.1 GI:1546071
SOURCE
Cochliobolus heterostrophus.
ORGANISM
Cochliobolus heterostrophus
Eukaryota; Fungi; Ascomycota; Euascomycetes; Loculoascomycetes;
Pleosporales; Pleosporaceae; Cochliobolus.
REFERENCE
1 (bases 1 to 9829)
Yang, G., Rose, M.R., Turgeon, B.G. and Yoder, O.C.
AUTHORS
Cochliobolus heterostrophus PKS1
TITLE
Plant Cell (1996) In press
JOURNAL
2 (bases 1 to 9829)
Yang, G., Rose, M.R., Turgeon, B.G. and Yoder, O.C.
AUTHORS
Direct Submission

BASE COUNT	68 a	40 c	52 g	53 t
ORIGIN				
3' UTR	1073 a	672 c	783 g	924 t
BASE COUNT				
ORIGIN				
Query Match	43.0%	Score 22.8;	DB 4;	Length 3452;
Best Local Similarity	92.3%	Pred. No. 25;		
Matches	24;	Conservative	0;	Mismatches 2; Indels 0; Gaps 0;
QY	10	AGGCACTCTAAGTCAATCATCC	35	
Db	260	AGGCATCTAGAGTCATCATTC	235	
RESULT 14				
AF155918/c	213 bp	DNA	ROD	16-JUN-1999
LOCUS	Sigmodon hispidus MHC class II antigen (Slh1-DQA) gene, Slh1-DQA*050			
DEFINITION	allele, partial cds.			
ACCESSION	AF155918			
NID	93070679			
VERSION	AF155918.1			
KEYWORDS	GI:5070679			
SOURCE	Sigmodon hispidus.			
ORGANISM	Sigmodon hispidus.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae; Sigmodon.			
AUTHORS	1 (bases 1 to 213)			
TITLE	Pfau, R.S., Van Den Busche, R.A., McBee, K. and Lochmiller, R.L.			
JOURNAL	Allelic diversity at the Mhc-DQA locus in cotton rats (Sigmodon hispidus) and a comparison of DQA sequences within the family Muridae (Mammalia: Rodentia)			
REFERENCE	Immunogenetics (1999) In press			
AUTHORS	2 (bases 1 to 213)			
TITLE	Pfau, R.S., Van Den Busche, R.A., McBee, K. and Lochmiller, R.L.			
JOURNAL	Submitted (02-JUN-1999) Zoology, Oklahoma State University, LSM 430, Stillwater, OK 74078, USA			
FEATURES	Location/Qualifiers			
source	1..213			
	/organism="Sigmodon hispidus"			
	/db_xref="taxon:42415"			
exon	1..213			
	/gene="Slh1-DQA"			
	/number=2			
gene	<1..>213			
	/gene="Slh1-DQA"			
	/allele="Slh1-DQA*05"			
mRNA	<1..>213			
	/gene="Slh1-DQA"			
	/product="MHC class II antigen"			
CDS	<1..>213			
	/gene="Slh1-DQA"			
	/codon_start=2			
	/product="MHC class II antigen"			
	/protein_id="AAD39248.1"			
	/db_xref="PID:95070680"			
	/db_xref="GI:5070680"			
	/translation="GMAIVYHSRKNQGYIFFDDDEKRYVDSKKETVWRIPEGGLT			
	SFDDQGGGLDATALAKYNLEIKRSNS"			

```

Query Match      43.0%: Score 22.8; DB 12; Length 213;
Best Local Similarity 66.0%: Pred. No. 26;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY      4 AAGCTTAGGCATTCTAAAGTCACATCATCCGACGTGAGTGCACATCGA 53
        ||| |||| | | | | | | | | | | | | | | | | | | | | |
Db       121 AAACCTCAGGAATCCTCCAGACAGTCTCTTTATCCGAGTCCACATAGA 72

RESULT 15
LOCUS    AF155922/c
DEFINITION
Sigmodon hispidus MHC class II antigen (Sih1-DQA) gene, Sih1-DQA*09 allele, partial cds.
ACCESSION
AF155922          213 bp      DNA           ROD         16-JUN-1999
Sigmodon hispidus MHC class II antigen (Sih1-DQA) gene, Sih1-DQA*09 allele, partial cds.
AF155922
AF5070687
AF155922.1   GI:5070687

SOURCE    Sigmodon hispidus.
KEYWORDS  Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
ORGANISM  Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae; Sigmodon.

REFERENCE
AUTHORS    1 (bases 1 to 213)
            Plau,R.S., Van Den Bussche,R.A., McBee,K. and Lochmiller,R.L.
TITLE       Allelic diversity at the Mhc-DQA locus in cotton rats (Sigmodon hispidus) and a comparison of DQA sequences within the family Muridae (Mammalia: Rodentia)
JOURNAL     Immunogenetics (1999) In press
REFERENCE   2 (bases 1 to 213)
AUTHORS    Plau,R.S., Van Den Bussche,R.A., McBee,K. and Lochmiller,R.L.
TITLE       Direct Submission
JOURNAL     Submitted (02-JUN-1999) Zoology, Oklahoma State University, LSU 430, Stillwater, OK 74078, USA

FEATURES
source     location/Qualifiers
            1..213 /organism="Sigmodon hispidus"
            1..213 /db_xref="taxon:42415"
            /gene="Sih1-DQA"
            /number=2
            <1..>213
            /gene="Sih1-DQA"
            /allele="Sih1-DQA*09"
            <1..>213
            /gene="Sih1-DQA"
            /product="MHC class II antigen"
            <1..>213
            /gene="Sih1-DQA"
            /codon_start=2
            /product="MHC class II antigen"
            /protein_id="AAD39252.1"
            /db_xref="PID:g5070688"
            /translation="GMALYISHPNGOYIFEPDGEKFIYDSDKKEVMIPIPEGLT
            SFDPGLSDIATVKHMLTLTKRSN"
BASE COUNT      66 a          42 c          52 g          53 t
ORIGIN

```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:21:48 ; Search time 133.08 Seconds
(without alignments)
99.641 Million cell updates/sec

Title: US-09-325-095-1718.SEO

Perfect score: 53
Sequence: 1 TTTAGCTTAGGCATTCTTAA.....CCGACTCGATCGACATCGA 53

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being pinned, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	66.0	35	1	Q59016
2	35	66.0	35	1	Q57022
3	26	49.1	3207	1	Q51155
4	26	49.1	3498	1	Q57012
5	24.4	46.0	3412	1	Q51156
6	24.4	45.3	1041	1	X15656
7	21.2	40.0	4130	1	V62749
8	20.8	39.2	11531	1	Q54222
9	20.8	39.2	6210	1	Q63815
10	20.8	39.2	53585	1	X20251
11	20.6	38.9	433	1	V64625
12	20.6	38.9	1906	1	V64625
13	20.2	38.1	2204	1	Q11675
14	20.2	38.1	2285	1	Q45935
15	20.2	38.1	2204	1	Q89532
16	19.8	37.4	225	1	Q31806
17	19.8	37.4	468	1	Q31811
18	19.8	37.4	314	1	Q77252
19	19.8	37.4	1628	1	Q74626
20	19.8	37.4	10669	1	V52190
21	19.6	37.0	1755	1	N91166
22	19.6	37.0	4648	1	Q12756
23	19.6	37.0	4648	1	Q12770
24	19.6	37.0	1196	1	T62680
25	19.6	37.0	1196	1	T62681
26	19.6	37.0	2089	1	V06591
27	19.6	37.0	11307	1	X12986
28	19.6	37.0	11000	1	X20248_01
29	19.4	36.6	7499	1	N91312
30	19.4	36.6	7824	1	Q74052
31	19.4	36.6	14690	1	X22303
32	19.4	36.6	10620	1	X02996
33	19.2	36.2	1423	1	O67209
34	19.2	36.2	1423	1	O67210
35	19.2	36.2	1421	1	O71698
36	19.2	36.2	133894	1	T13635
37	19.2	36.2	545	1	V53327
38	19.2	36.2	400	1	V78144
39	19.2	36.2	362	1	V78616
40	19.2	36.2	5137	1	V74528
41	19.2	36.2	13856	1	V74342
42	19.2	35.8	1714	1	N90718
43	19	35.8	1847	1	N90719

ALIGNMENTS

C 44	19	35.8	363	1	Q10877
C 45	19	35.8	5534	1	Q13574

RESULT 1	
ID	Q59016 standard; DNA: 35 BP.
AC	Q59016:
DT	12-APR-1994 (first entry)
DE	RACE PCR primer 2280.
KW	phosphoinositide kinase; PI: p85 subunit; screening; agonist;
KW	antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW	platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW	ss; amplification; p110.
OS	Synthetic.
PN	W09321528-A.
PD	28-OCT-1993.
PF	13-APR-1993; G00761.
PR	13-APR-1992; GB-008135.
PA	(LUDWIG) LUDWIG INST CANCER RES.
PI	Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
PI	Parkerp, Volinia S, Waterfield MD;
DR	WPI; 93-351738/44.
PT	Recombinant polypeptide(s) - with phosphoinositide-3 kinase
PS	activity, useful for controlling cell proliferation
CC	Example 1: Page 39; 146pp: English.
CC	An SGBAF-1 cell line was established by transfection of bovine
CC	adrenal cortex zona fasciculata cells with pSV3neo. Total RNA was
CC	isolated from the SGBAF-1 and a cDNA library constructed. RACE PCR
CC	was performed by synthesizing first strand cDNA from random hexamers
CC	on SGBAF-1 cell mRNA. PCR was performed using oligo 2224 and adaptor-
CC	dr as primers. Products were fractionated using an agarose gel. The
CC	cDNA was isolated from the gel and subjected to PCR using oligo 2280 and
CC	adaptor as primers. A product of 350 bp was further sequenced.
CC	See also Q51155-6, Q59012-23 and Q57522-3.
SO	Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;

Query Match	66.0%; Score 35; DB 1; Length 35;
Best Local Similarity	100.0%; Pred. No. 3.8e-06;
Matches	35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY	1 TTTAGCTTAGGCATTCTTAACTCATCATCC 35
DB	1 TTTAGCTTAGGCATTCTTAACTCATCATCC 35

RESULT 2	
ID	Q57022 standard; DNA: 35 BP.
AC	Q57022:
DT	31-AUG-1994 (first entry)
DE	Ptdins 3-Kinase oligo 2280
KW	110 kD catalytic subunit; phosphatidylinositol 3-kinase;
KW	transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW	ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
KW	blood vessel plaques; se.
OS	Synthetic.
PN	W09403609-A.
PD	17-FEB-1994.
PF	05-AUG-1993; G01651.
PR	05-AUG-1992; GB-016654.
PA	(IMCR) IMPERIAL CANCER RES TECHNOLOGY.
PI	Goode NT, Nurse PM, Parker PJ, Waterfield MD;
DR	WPI; 94-065697/08.
PT	Eukaryotic cells transformed with mammalian phospholipid or
PT	protein kinase DNA - useful in assays for compounds involved in
PT	cell growth regulation and for treating cancers
PS	Example 1: Page 15; 71pp: English.
CC	The sequences given in Q57020-23 are primers which were used in the

CC Isolation of the phosphatidylinositol (PtdIns) 3-kinase cDNA. The
 CC amplified sequence was placed under the regulatory control of the nmt
 CC promoter. These sequences were transformed into Schizosaccharomyces
 CC pombe cells in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the PtdIns
 CC plasmids grow as the parental strain. In the absence of thiamine the
 CC nmt promoter functions and the PtdIns is induced. PtdIns activity is
 CC substantially increased under these conditions. Cells containing
 CC constructs such as this, are useful in assays for detecting compounds
 CC involved in cell growth regulation. It is also used as the basis for
 CC detecting compounds for treating cancers and the formation of blood
 CC vessel plaques.
 SQ Sequence 35 BP; 10 A; 9 C; 4 G; 12 T;

Query Match 66.0%; Score 35; DB 1; Length 35;
 Best Local Similarity 100.0%; Pred. No. 3.8e-06;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTTAGCTTAGCATTTTAAAGTCATCATCC 35
 DB 1 TTTAGCTTAGCATTTTAAAGTCATCATCC 35

RESULT 3
 ID 051155 standard; cDNA; 3207 BP.
 AC 051155;
 DT 12-APR-1994 (first entry)
 DE p110 cDNA.
 KM Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KM platelets; neurophil activity; 3-phosphorylated phosphoinositides;
 KM ss.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"

PN MO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDWIG) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 PI WPI: 93-351738/44.
 DR P-PSDB: R43341.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Claim 7; Fig 9; 146pp; English
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC PI3 kinase p110 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p110 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with PI3 kinase activity, and is useful for
 CC screening for (ant)agonists of PI3 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051156, 059012-23 and 057522-3.
 SQ Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 49.1%; Score 26; DB 1; Length 3207;
 Best Local Similarity 100.0%; Pred. No. 0.075;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 10 AGGCATTCTAAAGTCATCATCC 35
 DB 10 AGGCATTCTAAAGTCATCATCC 35

DB 110 AGGCATTCTAAAGTCATCATCC 85

RESULT 4
 ID 057012/c
 DT 057012 standard; cDNA to mRNA; 3498 BP.
 AC 057012;
 DT 31-AUG-1994 (first entry)
 DE PtdIns 3-kinase 110 KD catalytic subunit cDNA.
 KM 110 KD catalytic subunit; phosphatidylinositol 3-kinase;
 KM transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
 KM PtdIns 3-kinase; assay; detection; cell growth; regulation; cancer;
 KM blood vessel plaques; ss.
 OS Bos taurus.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /product= p110

PN MO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 PI WPI: 94-065697/08.
 DR P-PSDB: R46294.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Disclosure: Fig 1; 71pp; English.
 CC This sequence encodes the 110 KD catalytic subunit of the phosphatidylinositol (PtdIns) 3-kinase. This sequence was transformed into
 CC Schizosaccharomyces pombe cells under the regulatory control of the
 CC nmt promoter in an embodiment of the invention. In the presence of
 CC thiamine the promoter is inactive and the cells carrying the PtdIns
 CC catalytic subunit plasmid grow as the parental strain. In the absence
 CC of thiamine the nmt promoter functions and the PtdIns 3-kinase
 CC catalytic subunit is induced. PtdIns activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SQ Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 49.1%; Score 26; DB 1; Length 3498;
 Best Local Similarity 100.0%; Pred. No. 0.076;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 AGGCATTCTAAAGTCATCATCC 35
 DB 110 AGGCATTCTAAAGTCATCATCC 85

RESULT 5
 ID 051156/c
 DT 051156 standard; cDNA; 3412 BP.
 AC 051156;
 DT 12-APR-1994 (first entry)
 DE Human p110 cDNA.
 KM Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KM antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KM platelets; neurophil activity; 3-phosphorylated phosphoinositides;
 KM ss.
 OS Human.
 FH Key Location/Qualifiers
 FT cds 1..3207
 FT /tag= a
 FT /note= "PI3- kinase p110"

PN MO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.
 PA (LUDM-) LUDMIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parkerp], Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 UK P-PSDB: RA33342.

PI Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PI activity, useful for controlling cell proliferation
 PS Claim 7; Fig 16; 146pp; English.
 CC Southern blot analysis was performed using a bovine cDNA probe contg.
 CC a fragment of a p13-kinase-encoding sequence and human cDNA isolated
 CC from a cDNA library constructed from mRNA isolated from the human
 CC cell line K562. Positive clones were sequenced to give the human
 CC p13 kinase p130 sequence shown. This sequence has 95 percent
 CC homology with the bovine sequence. The domain encoding residues 19-
 CC 100 of human p130 is sufficient to encode the kinase which will
 CC associate with the p85 kinase subunit. The gene may be used to
 CC provide a protein with p13 kinase activity, and is useful for
 CC screening for (ant)agonists of p13 kinase activity which could be
 CC useful for stimulation or inhibition of cell proliferation and hence
 CC for prophylaxis or therapy. Platelet or neutrophil activity or blood
 CC glucose levels can be controlled using the kinase.
 CC See also 051155 and 057522-3.
 SQ Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 46.0%; Score 24.4; DB 1; Length 3412;
 Best Local Similarity 96.2%; Pred. No. 0.36;
 Matches 25; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 10 AGGCATCTTAAGTCACTATCATCCC 35
 DB 110 AGGCATCTTAAGTCACTATCATTC 85

RESULT 6
 X15656/C
 ID X15656 standard; DNA: 1041 BP.
 AC X15656;
 DT 07-MAY-1999 (first entry)
 DE Protein phosphatase gene coding sequence of *Saccharomyces cerevisiae*.
 KW Protein phosphatase gene; growth; fermentation activity;
 KW dough production; yeast; ds.
 OS *Saccharomyces cerevisiae*.
 PN J11042090-A.
 PD 16-FEB-1999.
 PF 29-JUL-1997; 203652.
 PR 29-JUL-1997; JP-203652.
 PA (KANF) KANEKA CORP.
 PA (SHOS) SHOWA SANGYO CO.
 DR WPI: 99-197822/17.
 PI New yeast of controlled activation at low temperatures - useful for
 PI improving the quality of dough
 PS Claim 3; Page 19-20; 41pp; Japanese.
 CC The present sequence represents a protein phosphatase gene sequence
 CC of *Saccharomyces cerevisiae*. The specification describes new
 CC *S. cerevisiae* in which the growth and/or the fermentation activity
 CC is controlled at least in the range of 0-20 degrees Celsius. These
 CC yeast are prepared by deleting the function of at least one protein
 CC phosphatase gene. The yeast is useful in the production of dough.
 SQ Sequence 1041 BP; 553 A; 184 C; 243 G; 261 T;

Query Match 45.3%; Score 24; DB 1; Length 1041;
 Best Local Similarity 68.8%; Pred. No. 0.39;
 Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 4 AAGCTTAGCATTTAAAGTCACTATCATCCGAGCTGAGTCGACATC 51
 DB 756 AAGATCTGGCAATCCACCGCACTAGCACCTGACTGATGCATC 709

RESULT 7

V62749/C
 ID V62749 standard; CDNA: 4130 BP.
 AC V62749;
 DT 15-FEB-1999 (first entry)
 DE Human secreted protein clone er418.5 CDNA.
 KW Secreted protein; human; er418.5; ds.
 OS Homo sapiens.
 FT Key
 FT CDS Location/Qualifiers
 FT CDS 571..3306
 FT /*tag= a

PN W09846757-A2.
 PD 22-OCT-1998.
 PR 14-APR-1998; U07999.
 PR 13-APR-1998; US-059487.
 PR 15-APR-1997; US-843374.
 PA (GENM) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, LaValle ER, McCoy JM, Merberg D,
 PI Racie LA, Spaulding V, Treacy M;
 DR WPI: 98-568731/48.
 DR P-PSDB: W74723.

PI New polynucleotide(s) encoding secreted human proteins - are derived
 PI from, e.g. human foetal brain or foetal kidney cDNA libraries,
 PI potentially useful as, e.g. vaccines or thrombolytic agents
 PS Claim 19; Page 76-78; 120pp; English.
 CC Full-length cDNA clone er418.5 includes an open reading frame
 CC encoding a human secreted protein (see W74723). It was isolated
 CC from a human foetal brain cDNA library using methods which are
 CC selective for cDNAs encoding secreted proteins, or was identified
 CC as encoding a secreted or transmembrane protein on the basis of
 CC computer analysis of the amino acid sequence of the encoded protein.
 CC It shows some similarity to database sequences. The invention
 CC provides polynucleotides (see V62746-55) from human foetal brain,
 CC adult testis, adult brain, adult kidney and foetal kidney (all
 CC deposited as composite clone ATCC 98404), which encode human
 CC secreted proteins (see W74720-29). The polynucleotides and
 CC proteins are predicted to have biological activities which would
 CC make them suitable for treating, preventing or ameliorating medical
 CC conditions in humans and animals, although no supporting data are
 CC given. Suggested activities include nutritional activity, immune
 CC stimulating (e.g. as vaccines) or suppressing activity,
 CC haematopoiesis regulating activity, tissue growth activity,
 CC activin/inhibin activity, chemotactic/chemokinetic activity,
 CC haemostatic and thrombolytic activity, receptor/ligand activity,
 CC antiinflammatory activity, cadherin/tumour invasion suppressor
 CC activity, and tumour inhibition activity. The polynucleotides are
 CC also stated to be useful for gene therapy, and can be used in
 CC recombinant production of the polypeptides.
 SQ Sequence 4130 BP; 1113 A; 1051 C; 1024 G; 936 T;

Query Match 40.0%; Score 21.2; DB 1; Length 4130;
 Best Local Similarity 64.0%; Pred. No. 8.4;
 Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

OY 2 TTAAGCTTAGCATTTAAAGTCACTATCATCCGAGCTGAGTCGACATC 51
 DB 794 TTAATCTTAGCATTTTCAAGGAGACCTTCCTTCGTGTCAGACATC 745

RESULT 8
 O54222/C
 ID O54222 standard; DNA: 11531 BP.
 AC O54222;
 DT 27-JUN-1994 (first entry)

DE BSSU/CEL Gene.
 KW BSSU, biologically functional bile salt stimulated lipase; CEL;
 KW carboxyl ester lipase; hybridisation; milk protein; transgenic;
 KW infant milk substitute; oral; specific; lipolysis; vitamin; bile;
 KW lipid malabsorption; cystic fibrosis; chronic pancreatitis;
 OS digestion; ss.
 OS Homo sapiens.
 FT Key
 FT CDS Location/Qualifiers
 FT CDS 1653..11397

PN M0958943-A1.
PD 30-DEC-1998. U12764.
PE 18-JUN-1998; US-057483.
PR 03-SEP-1997; US-057483.
PR 20-JUN-1997; US-050359.
PR 22-JUL-1997; US-053344.
PR 22-JUL-1997; US-053377.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
PI Clayton R, Dougherty BA, Fraser C, Iathigra R, Smith HO,
PI White OR;
DR WPI: 99-081217/07.
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
PS Claim 1; Page 801-831; 1128pp; English.
CC X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (hb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T;

Query Match 39.2%; Score 20.8; DB 1; Length 53585;
Best Local Similarity 70.0%; Pred. No. 24;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 AGGATTCTAAAGTCACTATCATCCGAGCTCGAGTCGACA 49
44697 AAGCTTCAGAGAACCATCATCCACAGAGTCACACA 44658
|||||

RESULT 11
ID V31302 standard; DNA; 433 BP.
AC V31302;
NT 01-OCT-1998 (first entry)
DE E. coli J96 pathogenicity island contig #116.
KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pher;
KW PAI V; pher; vaccine; protective immune response; ds.
OS Escherichia coli.
FM WC3622575-A2.
PD 28-MAY-1998.
PE 21-NOV-1997; U21347.
PR 14-OCT-1997; US-061953.
PR 22-NOV-1996; US-031626.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYMI-) UNIV WISCONSIN.
PI Choi GH, Dillon PJ, Welch RA;
DR WPI: 98-312461/27.
PT New isolated uropathogenic E. coli nucleotide sequences - used to
PT develop products for the detection of pathogenic E. coli and to
PT elicit an immune response to pathogenic E. coli
PS Claim 21; Page 226; 250pp; English.
CC This sequence represents a E. coli strain J96 contig containing
CC pathogenicity island (PAI) sequences, and represents a nucleic acid
CC molecule of the invention. PAIs are large fragments of DNA which comprise
CC pathogenicity determinants. The sequences of the invention are taken from
CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pher)
CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at
CC approximately 94 min (at pher) on the E. coli chromosome and is
CC approximately 160 kb in size. Antibodies specific to the proteins encoded
CC by the PAI open reading frames of the invention can be used in kits to
CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit
CC a protective immune response in an animal to the uropathogenic E. coli
CC strain J96.
SQ Sequence 433 BP; 129 A; 94 C; 64 G; 145 T;

Query Match 38.9%; Score 20.6; DB 1; Length 433;
Best Local Similarity 62.7%; Pred. No. 8.4;
Matches 32; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 TTATAGCTTAGGCAATTAAGTCACTATCATCCGAGCTCGAGTCGACATC 51
289 TTACGCTTATACATGATGAGTCACTATCATCTTCGCTGGAAGGCCACATC 339
|||||

RESULT 12
ID V64625 standard; cDNA; 1906 BP.
AC V64625;
DR 02-MAR-1999 (first entry)
DE Tomato fructokinase (Frx1) cDNA.
KW Tomato; Polato; fructokinase; regulatory sequence; ripening; fruit;
KW carbohydrate metabolism; transgenic plant; expression cassette;
KW Lycopersicon esculentum; ss.
OS Lycopersicon esculentum.
FM Key Location/Qualifiers
FT CDS 271..1314
FT /*tag- a
FT /*note- "Fructokinase (frx1) coding region"

PN M09845412-A1.
PD 15-OCT-1998.
PE 10-MAR-1998; U04649.
PR 05-APR-1997; US-826611.
PA (REGC) UNIV CALIFORNIA.
PI Bennett AB, Kanayama Y;
DR WPI: 98-583193/49.
DR P-PSDB; W81785.
PT New fructokinase gene from tomato - useful for, e.g. inhibiting
PT fructokinase activity for increasing sweetness or stimulating starch
PT biosynthesis in plants
PS Claim 1; Page 28-32; 51pp; English.
CC The introduction of an expression cassette containing an exogenous Frx1
CC (fructokinase) sequence, is used to modify sweetness and promote starch
CC biosynthesis in transgenic plants, particularly tomato. Expression of
CC Frx1 in sense or antisense orientation inhibits conversion of fructose
CC to its 6-phosphate, increasing sweetness. Overexpression of
CC fructokinase is used to increase starch content in tubers, roots and
CC seeds.
SQ Sequence 1906 BP; 515 A; 362 C; 422 G; 607 T;

Query Match 38.9%; Score 20.6; DB 1; Length 1906;
Best Local Similarity 62.7%; Pred. No. 12;
Matches 32; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 TTATAGCTTAGGCAATTAAGTCACTATCATCCGAGCTCGAGTCGACATC 51
DB 1740 TTAAGTGCTGTATTATTAAGTCACTATCATGCTGTATTATTAATC 1790
|||||

RESULT 13
ID Q11675/c
AC Q11675;
DR 16-JUL-1991 (first entry)
DE H2RIIBP cDNA clone.
KW Major histocompatibility complex; MHC class I regulatory element;
KW nuclear hormone receptors; tumours; cancer; H2RIIBP; ss.
OS synthetic.
FM Key Location/Qualifiers
FT CDS 1..1341
FT /*tag- a
FT /*product- H2RIIBP protein
FT misc_rna 346..549
FT /*tag- b
FT /*label- zinc
FT poly_a_signal 1832..1837
FT /*tag- c

```

FT poly_a_signal 2185..2190
FT /tag- d
FT misc_rna 790..810
FT /tag- e
FT /label= glycosylation site
FT 826..852
FT /tag- f
FT /label= glycosylation site
PN US7450162-A.
PD 08-MAY-1990.
PE 13-DEC-1989; 163312.
PA (USSH ) NAT INST OF HEALTH.
PI Ozato K.
DR MPI: 91-140556/19.
DR P-PSDB: R11837.
PT New protein of nuclear hormone receptor super-family - used in
PT diagnosis and prognosis of treatment of tumours and other
PT conditions.
PS Disclosure: fig 3: 25pp: English.
CC This H2RIIB cDNA encodes a major histocompatibility complex
CC (MHC) class I regulatory element (CRE) binding factor. H2RIIB
CC binds to the wild-type MHC CRE region II sequence specifically.
CC It does not bind any other sequences except for the Vitellogenin A2
CC ERE sequence. It can be used for the diagnosis and prognosis of
CC tumours or conditions resulting from the imbalance in the normal
CC level of H2RIIB in cells or tissues.
CC See also Q11664-74.
SQ Sequence 2204 BP; 436 A; 645 C; 648 G; 475 T;

Query Match 38.1%; Score 20.2; DB 1; Length 2204;
Best Local Similarity 68.3%; Pred. No. 19;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

n 6 GCTTAGCATTCTAAGTCATCATCCGACTCGAGTCG 46
n 493 GCTTGCCACTGTACAGCTTTGTATACGACACGAGTAG 453

RESULT 14
v-v-v-v-v/c
AC 045955: standard; DNA; 2285 BP.
DE 01-FEB-1994 (first entry)
DE mrXR-beta.
KW Retinoid X receptor; RXR; retinoic acid receptor; RAR;
KW thyroid receptor; TR; retinoic acid response elements; RARE;
KW thyroid receptor response element; TRE;
KW retinoid X receptor response element; RXRE;
KW subunit; heterodimer; ss.
US Mus musculus.
FH Key Location/Qualifiers
FT cds 76..1422
FT /tag- a
FT /label= RXR-beta
PN MO9315216-A.
PD 05-AUG-1993.
PE 25-JAN-1993; U00639.
PA (CNRS ) CENT NAT RECH SCI.
PA (INRM ) INST NAT SANTE & RECH MEDICALE.
PA (SQU ) SQUIBB & SONS INC E R.
PA (U1ST-) UNITV PASTEUR STRASBOURG LOUIS.
PI Chambon P, Kastner P, Leid M;
DR P-PSDB: R39467.
PT Heterodimer comprising retinoic acid or thyroid receptor and
PT retinoid X receptor - useful in diagnosis and for generation of
PT antibodies
PS Claim 26; Page 84-86; 128pp: English.
CC Two different types of nuclear receptors, RAR and TR dimerise with
CC RXR to form a heterodimer which is capable of binding to RARE, TRE

```

```

CC or RXRE at physiological conditions. Sequences encoding 1 isoform
CC of mouse RXR-beta, 3 isoforms of human RXR-beta, 1 isoform of mouse
CC RXR-alpha and 1 isoform of mouse RXR-gamma are provided (045955-60).
SQ Sequence 2285 BP; 447 A; 665 C; 681 G; 492 T;

Query Match 38.1%; Score 20.2; DB 1; Length 2285;
Best Local Similarity 68.3%; Pred. No. 19;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 6 GCTTAGCATTCTAAGTCATCATCCGACTCGAGTCG 46
OY 574 GCTTGCCACTGTACAGCTTTGTATACGACACGAGTAG 534

Db 574 GCTTGCCACTGTACAGCTTTGTATACGACACGAGTAG 534

RESULT 15
ID 089532/c
ID 089532 standard; cDNA; 2204 BP.
AC 089532:
DE 24-OCT-1995 (first entry)
DE H-2RIIB cDNA.
KW H-2 region II binding protein; H-2RIIB; retinoid receptor; ss.
OS Not specified.
FH Key Location/Qualifiers
FT cds 1..1341
FT /tag- a
PN US5403925-A.
PD 04-APR-1995.
PE 13-DEC-1989; 450162.
PE 13-DEC-1989; US-450162.
PR 09-APR-1992; US-866950.
PR 28-SEP-1992; US-952800.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PI Ozato K.
DR MPI: 95-146856/19.
DR P-PSDB: R72484.
PT New DNA encoding human H-2 region II binding protein - an
PT intracellular retinoid receptor mediating effects of steroids)
PT etc.; also diagnostic kits contg. probes derived from the DNA
PS Disclosure: Column 19-24; 14pp: English.
CC A cDNA library of the Burkitt lymphoma cell line Daudi was screened
CC with a fragment of the murine retinoid X receptor-beta (H-2RIIB)
CC gene. A positive clone contained an insert (sequence given in
CC 089531) that encoded human H-2RIIB (R72483). Further H-2RIIB
CC sequences of unspecified origin (possibly murine) are given in
CC 089432 and R72484.
SQ Sequence 2204 BP; 436 A; 645 C; 648 G; 475 T;

Query Match 38.1%; Score 20.2; DB 1; Length 2204;
Best Local Similarity 68.3%; Pred. No. 19;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 6 GCTTAGCATTCTAAGTCATCATCCGACTCGAGTCG 46
OY 493 GCTTGCCACTGTACAGCTTTGTATACGACACGAGTAG 453

Db 493 GCTTGCCACTGTACAGCTTTGTATACGACACGAGTAG 453

```

Search completed: September 13, 1999, 16:21:51
Job time: 4011 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

us-09-325-095-1718 - nucleic search, using sw model

Run on: September 13, 1999, 16:13:52 ; Search time 849.94 Seconds

(without alignments)
123,002 Million cell updates/sec

Title: US-09-325-095-1718.SEQ

Sequence: 1 TTTAAGCTTAGCGATTCTAA.....CCGACTCGAGTCGACATCGA 53

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST:*

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	25	47.2	417	AA190409	AA190409 zp89c10.r
2	22.6	42.6	376	AI509866	AI509866 ms30c06.y
3	21.4	40.4	503	AA197781	AA197781 mw03f10.r
4	21.2	40.0	373	DA6325	DA6325 RICS10928A
5	21.2	40.0	275	AA357355	AA357355 EST60068
6	21.2	40.0	430	AA464819	AA464819 z443901.r
7	21	39.6	308	AA513147	AA513147 n129b07.s
8	21	39.6	492	AA987152	AA987152 uc80h09.x
9	20.8	39.2	200	C14053	C14053 C14053 Clon
10	20.8	39.2	404	AA525881	AA525881 n157f06.s
11	20.8	39.2	480	AA585790	AA585790 28439.Lam
12	20.8	39.2	307	AA610665	AA610665 np94a03.s
13	20.8	39.2	360	C61659	C61659 C61659 Yuj1
14	20.8	39.2	407	AA779495	AA779495 ef23b12.s
15	20.8	39.2	484	AA932095	AA932095 o062f01.s
16	20.8	39.2	350	AI025525	AI025525 ov75c10.x
17	20.6	38.9	445	W66698	W66698 me26c08.r1
18	20.6	38.9	594	AA438469	AA438469 LD10349.5
19	20.6	38.9	745	AA438912	AA438912 LD13306.5
20	20.6	38.9	504	AA392739	AA392739 LD11875.5
21	20.6	38.9	334	AA885288	AA885288 a158g04.s
22	20.6	38.9	520	AI104047	AI104047 EST213336
23	20.6	38.9	517	AI104124	AI104124 EST213413
24	20.6	38.9	530	AI105199	AI105199 EST214488
25	20.6	38.9	595	AI172362	AI172362 EST218366
26	20.6	38.9	487	AI232471	AI232471 EST229159
27	20.6	38.9	487	AA819739	AA819739 UT-R-A0-2
28	20.6	38.9	555	AI009655	AI009655 EST204106
29	20.6	38.9	599	AI511879	AI511879 S003164.5
30	20.6	38.9	471	AI722479	AI722479 fc27h07.x
31	20.4	38.5	356	R86883	R86883 Y930a05.r1
32	20.4	38.5	369	R89816	R89816 yp91e12.s1
33	20.4	38.5	366	W34996	W34996 mc34c02.r1
34	20.4	38.5	268	AA224491	AA224491 zt17d03.s
35	20.4	38.5	375	AA350887	AA350887 EST58398
36	20.4	38.5	561	AA518461	AA518461 v106h10.r
37	20.4	38.5	386	C25435	C25435 C25435 Rice
38	20.4	38.5	664	AA753536	AA753536 96BS0749
39	20.4	38.5	593	AA887564	AA887564 c141c12.s
40	20.4	38.5	817	AI037083	AI037083 ue19b08.y
41	20.4	38.5	609	AI0030452	AI0030452 A0030452
42	20.4	38.5	545	C97410	C97410 C97410 Rice
43	20.4	38.5	648	C97445	C97445 C97445 Rice
44	20.4	38.5	462	C97913	C97913 C97913 Rice
45	20.4	38.5	394	AU071270	AU071270 AU071270

ALIGNMENTS

RESULT 1
AA190409/c 417 bp mRNA EST 15-JAN-1997
LOCUS zp89c10.r1 Stratagene Hela cell s3 937216 Homo sapiens cDNA clone
DEFINITION IMAGE:627378 5' similar to SW-PI1A BOVIN P32871
ACCESSION PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPRA ISOPORM ;,
AA190409 mRNA sequence.

NID 91779240
 VERSION AAI90409.1 GI:1779240
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 1 (bases 1 to 417)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chappel, B., Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Marls, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rikkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevisan, E., Underwood, K., Woldmann, P., Waterston, R., Wilson, R. and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 On May 18, 1995 this sequence version replaced gi:811202.
 TITLE
 JOURNAL
 MEDLINE
 COMMENT
 Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 328.
 Location/Qualifiers
 1..417
 /organism="Homo sapiens"
 /db_xref="GDB:5046076"
 /db_xref="taxon:9606"
 /clone="IMAGE:627378"
 /clone.lib="Stratagene Hela cell s3 937216"
 /sex="female"
 /dev_stage="Hela S3 cell line"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Hela S3 epitheloid carcinoma cells grown to semi-confluency without induction. Average insert size: 1.5 kb; Uni-ZAP XR vector. -5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTT 3'."
 BASE COUNT 116 a 98 c 93 g 99 t 11 others
 ORIGIN
 Query Match 47.2%; Score 25; DB 29; Length 417;
 Best Local Similarity 75.6%; Pred. No. 2.7;
 Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Oy 10 AGGCATCTAAAGTCACTATCATCCGACGAGTCAGAT 50
 Db 292 AGGCATCTAAAGTCACTATCATCCATTTGGTAGAACAT 252
 RESULT 2
 LOCUS A1509866 376 bp mRNA EST 12-MAR-1999
 DEFINITION ms30e06.y1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:608482 5', mRNA sequence.
 ACCESSION A1509866
 NID 94408771
 VERSION A1509866.1 GI:4408771
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 376)

AUTHORS
 Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schuk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
 The WashU-NCI Mouse EST Project 1999
 Unpublished (1999)
 On Jun 5, 1998 this sequence version replaced gi:3188558.
 TITLE
 JOURNAL
 COMMENT
 Contact: Marra M/WashU-NCI Mouse EST Project 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 This read is a RESSEQUENCE of a previously sequenced mouse clone
 This read has been verified (found to hit its original self in the correct orientation)
 Seq primer: -40RP from Gibco
 High quality sequence stop: 375.
 Location/Qualifiers
 1..376
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /map="19p12-p13.1"
 /clone="IMAGE:608482"
 /clone.lib="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR vector; -5' adaptor sequence: 5' GAATTCGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTT 3'."
 BASE COUNT 106 a 97 c 80 g 93 t
 ORIGIN
 Query Match 42.6%; Score 22.6; DB 47; Length 376;
 Best Local Similarity 64.2%; Pred. No. 24;
 Matches 34; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
 Oy 1 TTTAAGCTTAAGTCACTATCATCCGACGAGTCAGATCGA 53
 Db 304 TTTAAGCTTAAGTCACTATCATCCGACGAGTCAGATCGA 356
 RESULT 3
 LOCUS AAI97781 503 bp mRNA EST 12-MAR-1997
 DEFINITION mv03f10.r1 Quaywoodford Beier mouse kidney day 0 Mus musculus cDNA
 clone IMAGE:653995 5', similar to gb:L47650 Mus musculus STRA6 mRNA,
 complete cds (MOUSE);, mRNA sequence.
 ACCESSION AAI97781
 NID 91793423
 VERSION AAI97781.1 GI:1793423
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 503)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.


```

Db      48  AAAGTTAAACATTCCAAAGCGACAACTTCTGTAATCACTGC 89

RESULT      6
LOCUS       AA464819/c
DEFINITION  zx49g01.r1 Soares total_fetus_Mb2HRF_9w Homo sapiens cDNA clone
IMAGE:789264 5', mRNA sequence.
ACCESSION   AA464819
NID         92189703
KEYWORDS    Homo sapiens
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE   1 (bases 1 to 430)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Getse1,G., Jost,S.,
            Kucba,T., Lacy,M., Le'N., Lennon,G., Matra,M., Martin,J.,
            Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
            WASHU-Merck EST Project 1997
            Unpublished (1997)
            On May 5, 1995 this sequence version replaced gi:788051.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
This clone is available royalty-free through LNLN; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28m3 rev2 ET from Amersham
High quality sequence stop: 417.
Location/Qualifiers
    1..430
    /organism="Homo sapiens"
    /db_xref="GDB:5986881"
    /db_xref="taxon:9606"
    /map="6"
    /clone="IMAGE:789264"
    /clone.lib="Soares total_fetus_Mb2HRF_9w"
    /dev_stage="8-9 weeks"
    /lab_host="DH10B"
    /note="vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDN
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dt) primer [5
TGTTCGAATCGAGTCGAGCGACGCCGCTTAATTTTCTTTTCTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
and Eco RI sites of the modified pRT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT      130 a      83 c      102 g      115 t
ORIGIN

Query Match      40.0%; Score 21.2; DB 34; Length 430;
Best Local Similarity 64.0%; Pred. No. 86;
Matches 32; Conservative 0; Mismatches 18; Indels 0; Gaps 0

cy      1  TTTAAGCTTAGCATCTCTAAAGTCATATCATCCGACATCGAGTCGACAT 50
        |||||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
cl      350  TTTAAGTTTGGCAGCCATATGCTCTGTCATGACATCAAACTCTTCAT 301

RESULT      7
LOCUS       AA513147
DEFINITION  n129b07.s1 NCI_CGAP_Ew1 Homo sapiens cDNA clone IMAGE:971701, mRNA
sequence.
AA513147/c      308 bp      mRNA      EST      13-NOV-1997

```

[illegible]


```

/lab_host="DH10B"
/Note="Vector: PAMPI0; mRNA made from invasive ovarian
tumor; cDNA made by oligo-dt priming. Non-directionally
cloned. Size selected on agarose gel, average insert size
600 bp. Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT      110 a      90 c      123 g      81 t
ORIGIN

Query Match
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 AGGCATTCTAAGTCACTATCATCCGACTCGATCGACA 49
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 AGCAATCCCAATTCATTAGATCAGGACTTGATTACA 25

RESULT 11
AA585790 480 bp mRNA EST 11-SEP-1997
LOCUS 28439 Lambda-PRL2 Arabidopsis thaliana cDNA clone 105J12XP 3', mRNA
DEFINITION
Sequence.
ACCESSION AA585790
NID 92393202
VERSION AA585790.1 GI:2393202
KEYWORDS EST.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M.,
Retzel, E. and Somerville, C.
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsids.
1 (bases 1 to 480)
Neuman, T., deBruin, F. J., Green, P., Keegstra, K., Kende, H.,
Arabidopsids.
Arabidopsids.
TITLE Genes galore: a summary of methods for accessing results from
JOURNAL large-scale partial sequencing of anonymous Arabidopsis cDNA clones
MEDLINE Plant Physiol. 106, 1241-1255 (1994)
COMMENT 95148729
On May 5, 1995 this sequence version replaced gi:798282.

Contact: Thomas Neuman
MSU-DOE Plant Research Laboratory
Michigan State University
MSU-DOE-PRL, Michigan State University, Plant Biology Bldg., E.
Lansing, MI
Tel: 517-353-0854
Fax: 517-353-9168
Email: 22313tcn@lhm.cl.msu.edu
The sequence entry for this EST has been reverse complemented and
is being submitted in the sense orientation.
Seq primer: M13 -21 dye primer.
Location/Qualifiers
1. 480
/organism="Arabidopsis thaliana"
/strain="var columbia"
/db_xref="taxon:3702"
/map="20"
/clone="105J12XP"
/clone_1lb="Lambda-PRL2"
/Note="Vector: lambda Z1p-Lox; Site_1: Sal; Site_2: Not;
Lambda PRL2 is a cDNA library derived from equal
quantities of 4 pools of mRNA. The mRNA sources were 1) 7
day germinated etiolated seedlings; 2) tissue culture
grown roots; 3) staged plants half with 24 hour light
cycle, half on 16 hr light, 8 hour dark - rosettes; 4)
same plants as 3 but aerial tissue (stems, flowers and
siliques). The vector is PRL's lambda Z1p-Lox. The cDNA
inserts were directionally cloned with Sal-Not arms using
oligo dt primed cDNA."

```

```

BASE COUNT      146 a      79 c      102 g      142 t      11 others
ORIGIN

Query Match
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 10 AGGCATTCTAAGTCACTATCATCCGACTCGATCGACA 49
      ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 230 AGTGGTCTTAAGTCTCCATGACACTACTCGATCGACACA 191

RESULT 12
AA610665 307 bp mRNA EST 09-DEC-1997
LOCUS np94a03.s1 NCI CGAP Thy1 Homo sapiens cDNA clone IMAGE:1133932
DEFINITION similar to TR:G1236142 G1236142 MARINER TRANSPOSASE; contains Alu
repetitive element; contains element LI repetitive element ; mRNA
sequence.
ACCESSION AA610665
NID 92459093
VERSION AA610665.1 GI:2459093
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 307)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL Tumor Gene Index
COMMENT Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1394357.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNM at:
www.bio.liml.gov/bbrp/image/image.html
Insert Length: 543 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 297.
Location/Qualifiers
1. 307
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1133932"
/clone_1lb="NCI CGAP Thy1"
/tissue_type="thyroid"
/lab_host="DH10B"
/Note="Vector: PAMPI0; mRNA made from invasive thyroid
tumor, cDNA made by oligo-dt priming. Non-directionally
cloned. Size selected on agarose gel, average insert size
600 bp. Reference: Krizman et al. (1996) Cancer Research
56:5380-5383."
BASE COUNT      85 a      70 c      66 g      86 t
ORIGIN

Query Match
Best Local Similarity 78.1%; Pred. No. 1.2e+02;
Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 2 TTAAGCTTAGGATCTTAAGTCACTATCATC 33
      . . . . . ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 71 TAAAGTTGGCTTCTCAAGACTCTCATC 40

RESULT 13

LOCUS C61659

DEFINITION C61659 Yui1 Kohara unpublished cDNA Caenorhabditis elegans CDNA clone yk2447 5', mRNA sequence.

ACCESSION C61659

VERSION 92420364

KEYWORDS C61659.1 GI:2420364

SOURCE EST.

ORGANISM Caenorhabditis elegans.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

AUTHORS 1 (bases 1 to 360)

* Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Mishigaki, A.

TITLE Expression map of the C.elegans genome

JOURNAL Unpublished (1996)

COMMENT On May 9, 1995 this sequence version replaced gi:802313.

FEATURES

source

Contact: Yui1 Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykharad@dbj.nig.ac.jp.

Location/Qualifiers

1..360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db_xref="taxon:6239"

/clone="yk2447"

/clone_lib="Yui1 Kohara unpublished cDNA"

/note="dev_stage=varied, sex=Hermaphrodite male, tissue_type=whole animal"

BASE COUNT 104 a 78 c 77 g 101 t

ORIGIN

Query Match 39.2%; Score 20.8; DB 36; Length 360;
Best Local Similarity 70.0%; Pred. No. 1.2e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 14 ATCTAAAGTCACTATCATCCGACCTCGACATCGA 53

Db 152 ATCTAAAGTCACTATCATCCGACCTCGACATGGA 191

RESULT 14

LOCUS AA779495

DEFINITION AA779495 407 bp mRNA EST 05-FEB-1998
a123b12.s1 Soares-total_fetus_Nb2HF8_9w Homo sapiens CDNA clone
IMAGE:1032479 3', mRNA sequence.

ACCESSION AA779495

NID 92838826

VERSION AA779495.1 GI:2838826

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS 1 (bases 1 to 407)

Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,

Kirman, D., Kucada, T., Lacy, M., Le, N., Lennon, G., Marra, M.,

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,

Thesing, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

WashU-NCI human EST Project

TITLE Unpublished (1997)

COMMENT On Jan 19, 1998 this sequence version replaced gi:2287157.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: estevenson.wustl.edu

This clone is available royalty-free through LINT; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40m13 fwd. ET from Amersham

High quality sequence stop: 369.

FEATURES

source

1..407

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1032479"

/clone_lib="Soares-total_fetus_Nb2HF8_9w"

/dev_stage="8-9 weeks"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand CDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5']
TGTACCAATCTGAGTGGAGCGCGCTTAATTTTTTTTTTTT 3'].

Double-stranded CDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 85 a 95 c 123 g 104 t

ORIGIN

Query Match 39.2%; Score 20.8; DB 38; Length 407;
Best Local Similarity 70.0%; Pred. No. 1.3e+02;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 3 TAAGCTTGGCACTTAAAGTCACTATCATCCGACCTCGA 42

Db 254 TAAGCAAGCAAGCTCAAGACAGATCTCCAGACACCA 215

RESULT 15

LOCUS AA932095

DEFINITION AA932095 484 bp mRNA EST 07-JUL-1998
0062f01.s1 NCI-CGAP JUS Homo sapiens CDNA clone IMAGE:1570777 3'
similar to gp:M35718 FIBROBLAST GROWTH FACTOR RECEPTOR BFR-2
PRECURSOR (HUMAN);, mRNA sequence.

ACCESSION AA932095

NID 93086408

VERSION AA932095.1 GI:3086408

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

AUTHORS 1 (bases 1 to 484)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP).

Tumor Gene Index

Unpublished (1997).

JOURNAL

COMMENT On Sep 12, 1996 this sequence version replaced gi:1397953.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/BLN at:

www.bio.liml.gov/bbrp/image/image.html

Insert Length: 794 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 252.

FEATURES

source

1. .484

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_image="1570777"

/clone_lib="NCI CGAP L45"

/issue_type="carcinoid"

/lab_host="DH10B"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; 1st strand cDNA was prepared from neuroendocrine lung carcinoid, and was then primed with a Not I - oligo(dt) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 165 a 104 c 89 g 126 t

ORIGIN

Query Match 39.28; Score 20.8; DB 40; Length 484;
Best Local Similarity 64.6%; Pred. No. 1.3e+02;
Matches 31; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 4 AACGTTAGCATCTCTAAGTCACATCATCCGACGAGTGCACATC 51
||| | || |||| | |||| | |||| | |||| | |||| | ||||
Db 116 AAGGTAGTTATTCACATTCACCACTCCCGACTGTGACCCGCCACC 163

Search completed: September 13, 1999, 16:13:55
Job time: 5454 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:36 ; Search time 329.34 Seconds

(without alignments)
386.265 Million cell updates/sec

Title: US-09-325-095-2122.SEQ

Perfect score: 40

Sequence: 1 TGCTGTAATCTTAATGCTGGATTTCATGAACAAATGA 40

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*
1: gb_bal:*
2: gb_baz:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vl:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_or:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vl:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_bal:*
39: em_baz:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	26	65.0	185092	34	HSJ661G12
					AL049797 Homo sapi

C 2	23.8	59.5	28065	11	AC005354	AC005354 Homo sapi
C 3	23.2	58.0	86765	11	AC004472	AC004472 Homo sapi
C 4	23	57.5	110102	7	ATF28J12	AL021110 Arabidops
C 5	23	57.5	161652	11	AC004774	AC004774 Homo sapi
C 6	23	57.5	103911	11	HUAC002565	AC002565 Human Chr
C 7	22.8	57.0	18227	12	AF110420	AF110420 Mus muscu
C 8	22.4	56.0	109160	8	AC002328	AC002328 genomic s
C 9	22.4	56.0	110858	10	HUM263A	AC000062 Human PAC
C 10	22.4	56.0	170215	34	AC004127	AC004127 Homo sapi
C 11	22.4	56.0	200965	34	AC006740	AC006740 Caenorhab
C 12	22.4	56.0	37678	36	CEIK08B12	U97001 Caenorhabd1
C 13	22.2	55.5	236496	34	AC006790	AC006790 Caenorhab
C 14	22.2	55.5	19566	37	CELI49F8C	AF123232 Sequence 1
C 15	22	55.0	3498	5	A37232	AP000084 Homo sapi
C 16	22	55.0	100000	9	AP000084	AP000136 Homo sapi
C 17	22	55.0	100000	9	AP000136	AP000221 Homo sapi
C 18	22	55.0	89323	9	AP000221	AC006765 Caenorhab
C 19	22	55.0	274498	34	AC006765	Continuation (5 of
C 20	22	55.0	110000	34	CEY11B2_4	Continuation (2 of
C 21	22	55.0	110000	34	CEY37D8_1	298870 Caenorhabd1
C 22	22	55.0	291692	34	CEY59A8	AC005978 Drosophi1
C 23	22	55.0	85316	35	AC005978	292775 Caenorhabd1
C 24	22	55.0	41299	36	CEC06H5	AL032626 Caenorhab
C 25	22	55.0	132742	36	CEY37D8A	AF077539 Caenorhab
C 26	22	55.0	16938	37	CELT7SD3	AC006257 Homo sapi
C 27	21.8	54.5	175818	11	AC006257	U67483 Methanococ
C 28	21.6	54.0	9639	2	U67483	AC005292 Genomic s
C 29	21.6	54.0	99053	8	AC005292	AC004083 Homo sapi
C 30	21.6	54.0	177726	11	AC004083	AF042895 Homo sapi
C 31	21.6	54.0	330250	11	AF049895	298862 Caenorhabd1
C 32	21.6	54.0	43888	34	CEY43E12	281454 Caenorhabd1
C 33	21.6	54.0	29731	36	CEB0391	268748 Caenorhabd1
C 34	21.6	54.0	32320	36	CEY37H10	U24116 Hydra vulga
C 35	21.6	54.0	2732	36	HVU24116	AL034560 Plasmodiu
C 36	21.6	54.0	108908	36	PEMA13P8	AF063234 Pneumocys
C 37	21.4	53.5	3484	8	AE063234	AL008654 Human DNA
C 38	21.4	53.5	152992	9	HS396D17	Z84480 Homo sapien
C 39	21.4	53.5	121424	9	HS418A9	AC003964 Homo sapi
C 40	21.4	53.5	117435	11	HUAC003964	AC007297 Homo sapi
C 41	21.4	53.5	226683	35	AC007297	AF056942 Caenorhab
C 42	21.4	53.5	40114	36	CEL2K6	Z99531 S.pombe chr
C 43	21.2	53.0	30389	7	SPAC19D5	Z98981 S.pombe chr
C 44	21.2	53.0	44092	7	SPAC6F6	AL021152 Caenorhab
C 45	21.2	53.0	110000	34	CEY81G3_0	

ALIGNMENTS

RESULT 1
HSJ661G12/c
LOCUS Homo sapiens chromosome 1 clone J661G12, WORKING DRAFT SEQUENCE, 1n
DEFINITION Homo sapiens chromosome 1 clone J661G12, WORKING DRAFT SEQUENCE, 1n
UNORDERED PIECES.
ACCESSION AL049797
NID G5051384
VERSION AL049797.2 GI:5051384
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 185092)
AUTHORS Frankland, J.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: d661g12 Contig_ID: 02987 acc-AL049797 Length: 123411 bp Unfinished: d661g12 Contig_ID: 02328 acc-AL049797 Length: 15975 bp Unfinished: d661g12 Contig_ID: 00952 acc-AL049797 Length: 44106 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Location/Qualifiers

1. 185092

/db_xref="taxon:9606"

/db_xref="taxon:9606"

/db_xref="taxon:9606"

BASE COUNT 56011 a 32549 c 34225 g 60707 t 1600 others

ORIGIN

Query Match 65.0%; Score 26; DB 34; Length 185092;

Best Local Similarity 85.3%; Pred. No. 10;

Matches 29; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 AAATCTAATGCTGATTCATGAACAATGA 40

Db 169799 AAATCTAATGCTGATTCATGAACAATGA 169766

RESULT 2 AC005354 28065 bp DNA PRI 31-JUL-1998

LOCUS AC005354 AC002244 AC001059 AC001060 AC001536 L81878 L81879

DEFINITION Homo sapiens chromosome 5, P1 clone 1320c6 (UBNL H59), complete

sequence.

AC005354 AC002244 AC001059 AC001060 AC001536 L81878 L81879

L81880 L81881

93366558

AC005354.1 GI:3366558

HTG.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 28065)

Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Kadner, K., Miguel, T., Miller, C., Plick, S., Pollard, M.,

Rojas, H., Subramanian, S. and Martin, C.H.

Sequencing of human chromosome 5

Unpublished

2 (bases 1 to 28065)

Ricke, D.O.

Large Scale Sequence Analysis and Annotation with the Sequence

Comparison Analysis (SCAN) System

Unpublished

3 (bases 1 to 28065)

Kimmerly, W., Bondoc, M., Cheng, J., Connolly, K.S., Gunning, K.M.,

Davis, C.A., Kadner, K., Miguel, T., Plick, S., Pollard, M.,

Rojas, H., Subramanian, S. and Martin, C.H.

Submitted (31-JUL-1998) Human Genome Center, DOE Joint Genome

Institute, Lawrence Berkeley National Laboratory, MS 74-157,

Berkeley, CA 94720, U.S.A.

Sequence submitted by:

DOE Joint Genome Institute.

Location/Qualifiers

1. 28065

/db_xref="taxon:9606"

/db_xref="taxon:9606"

/db_xref="taxon:9606"

/db_xref="taxon:9606"

/db_xref="taxon:9606"

/db_xref="taxon:9606"

/db_xref="taxon:9606"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"

repeat_region /rpt_family="L1"


```

repeat_region 6113. .6662
/rpt_family="MER41A"
repeat_region complement(6663. .6994)
/rpt_family="MER83"
repeat_region complement(7000. .7557)
/rpt_family="MER4_internal"
repeat_region complement(7560. .8234)
/rpt_family="MER4A"
repeat_region 8239. .9165
/rpt_family="MSTB-internal"
repeat_region complement(9215. .9483)
/rpt_family="AluUo"
repeat_region 9484. .9748
/rpt_family="MSTB-internal"
repeat_region 9841. .10206
/rpt_family="MSTB"
repeat_region complement(10291. .10484)
/rpt_family="AluUo"
repeat_region complement(10495. .10794)
/rpt_family="AluSg"
repeat_region 11777. .12076
/rpt_family="AluSx"
repeat_region 12077. .12105
/rpt_family="AluSx"
repeat_region 12479. .12775
/rpt_family="TAAAn"
repeat_region 13479. .13511
/rpt_family="AluY"
repeat_region complement(13041. .13531)
/rpt_family="LINE2"
repeat_region 13809. .13944
/rpt_family="AluSx"
repeat_region 13945. .14247
/rpt_family="AluSg"
repeat_region 14248. .14417
/rpt_family="AluSx"
repeat_region 14504. .14618
/rpt_family="AluSg1"
repeat_region 14620. .14895
/rpt_family="AluSg1"
repeat_region 14896. .15064
/rpt_family="AluSg"
repeat_region 15118. .15404
/rpt_family="AluSg1"
repeat_region 15334. .15566
/rpt_family="AluSg"
misc_feature 15334. .15566
/note="DGS similarity to overlapping ESTs:-(15334. .15694)
AA013249 ze26c07.s1 Soares retina N2b4HR Homo sapiens cDNA
clone 360108 3' similar to gb:M35663 INTERFERON-INDUCED,
DOUBLE-STRANDED RNA-ACTIVATED PROTEIN KINASE
(HUMAN):contains Alu repetitive element; Score: 645
identity: 359/370 (97%)--(15334. .15566) AA016313
ze38h03.s1 Soares retina N2b4HR Homo sapiens cDNA clone
361301 3' similar to gb:M35663 INTERFERON-INDUCED,
DOUBLE-STRANDED RNA-ACTIVATED PROTEIN KINASE
(HUMAN):contains Alu repetitive element; Score: 449
identity: 230/232 (99%)."
repeat_region 16868. .16930
/rpt_family="MLT1A2"
repeat_region 16932. .16978
/rpt_family="POLY-A"
repeat_region complement(17033. .17060)
/rpt_family="AT-rich"
misc_feature 17101. .17239
/note="DGS similarity to AA054173 z551a08.r1 Soares retina
N2b4HR Homo sapiens cDNA clone 380438 5' similar to
contains Alu repetitive element; (416. .277); 99%
identity."
repeat_region 17188. .17319
/rpt_family="AluUo"
repeat_region 17540. .17797
/rpt_family="AluSg"
misc_feature 17714. .18004
/note="DGS similarity to overlapping ESTs:-(17728. .18004)
AA054173 z551a08.r1 Soares retina N2b4HR Homo sapiens cDNA
clone 380438 5' similar to contains Alu repetitive

```

```

element; (276. .1); 99% identity.-(18004. .17714) AA013121
ze26c07.r1 Soares retina N2b4HR Homo sapiens cDNA clone
360108 5' similar to contains Alu repetitive element;
(1. .292); 97% identity.-(18004. .17714) AA017502
ze38h03.r1 Soares retina N2b4HR Homo sapiens cDNA clone
361301 5' similar to contains Alu repetitive element;
Score: 567 identity: 292/297 (98%)."
repeat_region 17892. .18191
/rpt_family="AluUo"
repeat_region complement(18516. .18710)
/rpt_family="AluUo"
repeat_region 18906. .19203
/rpt_family="AluY"
repeat_region 19288. .20021
/rpt_family="LINE2"
repeat_region 20258. .20313
/rpt_family="LINE2"
repeat_region complement(20739. .21323)
/rpt_family="IRRI0C"
repeat_region complement(21334. .21361)
/rpt_family="MIR"
repeat_region 21365. .21487
/rpt_family="FLAM_A"
repeat_region 21491. .21793
/rpt_family="AluUo"
repeat_region complement(21817. .21953)
/rpt_family="MIR"
misc_feature 22765. .22991
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 86.000"
repeat_region 22810. .22867
/rpt_family="GAAAn"
misc_feature 24453. .24713
/note="DGS similarity to A362319 EST95595 Testis I Homo
sapiens cDNA 5' end. Score: 505 identity: 258/260 (99%)."
misc_feature 25318. .25887
/note="DGS similarity to AA061824 m133d04.r1 Stratagene
mouse testis (#937308) Mus musculus cDNA clone 513799 5';
Score: 279 identity: 295/437 (67%)--(25480. .25887)
BLASTN similarity to AA860210 (1. .407); match: 1, score:
5.7e-157; database searched: month.na; ak48c12.s1 Soares
testis NHT Homo sapiens cDNA clone IMAGE:1409206 3'"
repeat_region 26514. .26804
/rpt_family="MLT1D"
repeat_region complement(26831. .27035)
/rpt_family="AluUo"
repeat_region complement(27042. .27333)
/rpt_family="AluY"
repeat_region complement(27334. .27637)
/rpt_family="AluSp"
repeat_region complement(27644. .27716)
/rpt_family="MLT1D"
repeat_region 27727. .28020
/rpt_family="AluSx"
repeat_region 28681. .28972
/rpt_family="AluSx"
repeat_region 29177. .29471
/rpt_family="7SLRNA"
misc_feature complement(29214. .29339)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: good, score: 50.000"
Query Match 58.0%; Score 23.2; DB 11; Length 86765;
Best Local Similarity 77.8%; Pred. No. 87;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY 1 TGCTGTAATTCATGCTGATTCATGACAA 36
11706 TGCTGTAATTCATGCTGATTCATGACAA 11671

```



```

exon      8557..8838
          /gene="F28J12.30"
          /number=1
gene       8557..13947
          /gene="F28J12.30"
          /join(8557..8838,8913..9050,9136..9198,9284..9373,
          9450..9572,9668..9755,10089..10163,10281..10280,
          10348..10368,12117..12725,12806..13321,13362..13622,
          13693..13947)
          /gene="F28J12.30"
          /note="similarity to MucD - Pseudomonas aeruginosa,
          PATCH:G1184684; contains EST gb:722208"
          /codon_start=1
          /product="putative protein"
          /protein_id="CA16717.1"
          /db_xref="PID:1249602"
          /db_xref="PID:9283642"
          /db_xref="GI:2832642"
          /db_xref="SPTREMBL:O49507"
          /translation="MALASSKSSIFNTLSPINOSKPYLACSGSNHYDIDRRRI
          IFGSLATSLGSGNOQRLEPESHIALDQFEKEEELFEERNVNLFOKTPSPVY
          IEALPRTSSGDLTDEENGLEGSGFVMDKLGHVTVNHVIAKLATDQFGQRC
          KSLVDANGTRFSKEGKIVGLDNDLAVLKLEEGREINPVYLGTSNDLRVQSCFA
          IGNPYGENTLIGVYSGIAGREIPSPNGKSISEALQTDADINSNGSGILDYGHFI
          GVTATFERKSGSSGLIETDSEVLNSVYILMVERKKRKROIORNSNMRKRIS
          HDKINRDELIVYRIICPIDVGVGIGSGKVNAIRHTKAKIRFQDLHGCSQVIT
          IYCSYKROEEIGFTKSENEPCCAQDALIKYDAIVASDENNTKTVDRDNECR
          LVPFSQSSSLIGKAGENIKRLRRRTKASVKVSKVSDPSHVCAEMEDNVVISGE
          ESKQALFAVSAIMKINRENIPIDSTSDVPASVIVPSDLSNVSYPQGFYGNOD
          HLOQAGAVPSYFNAISVDEFGVAETANPVPAASIVPFGHSGSSRSEELFVKY
          LCPICNIMHVIGSGSTIKRIRASGSCIEVNSDKGDDCVIIVATLIFCCIS
          TPFEVMSPPDDKMSMAVEAVILLOEYINDEDEENVKQMLYASGVIGSGSVIN
          ETRKTNANICISKRKDDLYVESEVSSVRDALQIYLRLEDVIGDKDSVATRKPP
          ARTDNCSTLSSGSSNAGITLPSPMSSMASTSGFHVGSPPVSCILDFH"
          8839..8912
          /gene="F28J12.30"
          /number=1
exon      8913..9050
          /gene="F28J12.30"
          /number=2
          9051..9135
          /gene="F28J12.30"
          /number=2
          9136..9198
          /gene="F28J12.30"
          /number=3
          9199..9283
          /gene="F28J12.30"
          /number=3
          9284..9373
          /gene="F28J12.30"
          /number=4
          9374..9449
          /gene="F28J12.30"
          /number=4
          9450..9572

```

Query Match Best Local Similarity 57.5%; Score 23; DB 7; Length 110102;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

```

0Y 1 TGCTGTAATTCATGCTGATTTCATGAACAATG 39
DB 80490 TGATGTAAATTTAATTGATATTCTCGTAAATTAATG 80452

```

```

RESULT 5
AC004774 161652 bp DNA PRI 01-OCT-1998
LOCUS AC004774 Homo sapiens BAC clone RG300E22 from 7q21-q31.1, complete sequence.
ACCESSION AC004774
NID 93169209
VERSION AC004774.1 GI:3169209

```

KEYWORDS SOURCE ORGANISM REFERENCE AUTHORS TITLE JOURNAL REFERENCE AUTHORS TITLE JOURNAL COMMENT

HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 161652)
Zhu, H., Hinde, K. and Strommat, C.
The sequence of Homo sapiens BAC clone RG300E22
Unpublished (1998)
2 (bases 1 to 161652)
Waterson, R.
Direct Submission
Submitted (30-MAY-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
SUBMITTED BY: WUGSC
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
Clone RG300E22 is from a release of the human BAC library CITB-HS-A. The library contains cloned DNA from human sperm. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). The clone is available from Research Genetics, Inc. (<http://www.resgen.com>).
VECTOR: pBelobAC11
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:
The actual start of this clone is at base position 1 of RG300E22; actual end is at 161652 of RG300E22.

This clone contains STS's SHGC-15930 (NID:91132038) and SWS3387 (NID:9339150).

FEATURES
source
Location/Qualifiers
1..161652
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-q31.1"
/clone_1jb="CITB-HS-A"
/clone="RG300E22"
3089..3265
/rpt_family="MALR"
3296..3476
/rpt_family="(TAGA)n"
3494..3677
/rpt_family="MALR"

```

repeat_region      7510..7596
                    /rpt_family="MERL_type"
repeat_region      7809..7910
                    /rpt_family="L2"
repeat_region      8148..8256
                    /rpt_family="L2"
repeat_region      8480..8667
                    /rpt_family="L1"
repeat_region      9112..9205
                    /rpt_family="MERL_type"
repeat_region      11511..11804
                    /rpt_family="Alu"
repeat_region      13624..13993
                    /rpt_family="MALR"
repeat_region      14122..14252
                    /rpt_family="MIR"
repeat_region      20744..21164
                    /rpt_family="L1"
repeat_region      21166..21314
                    /rpt_family="L1"
repeat_region      21331..21614
                    /rpt_family="Alu"
repeat_region      21615..21664
                    /rpt_family="(CATR)n"
repeat_region      21665..22121
                    /rpt_family="L1"
repeat_region      22171..22237
                    /rpt_family="MIR"
repeat_region      23047..23347
                    /rpt_family="Alu"
repeat_region      23357..23718
                    /rpt_family="MALR"
repeat_region      24659..24960
                    /rpt_family="Alu"
repeat_region      25327..25559
                    /rpt_family="L1"
repeat_region      25826..25949
                    /rpt_family="L1"
repeat_region      29100..29380
                    /rpt_family="Alu"
repeat_region      29385..29482
                    /rpt_family="(TAGA)n"
repeat_region      29544..29646
                    /rpt_family="L2"
repeat_region      30131..30242
                    /rpt_family="L2"
repeat_region      30355..30653
                    /rpt_family="Alu"
repeat_region      30948..31107
                    /rpt_family="MERL_type"
repeat_region      33110..34309
                    /rpt_family="MERL_type"
repeat_region      34416..34739
                    /rpt_family="L2"
repeat_region      35369..35571
                    /rpt_family="L2"
repeat_region      35364..35798
                    /rpt_family="L2"
repeat_region      35818..36225
                    /rpt_family="L2"
repeat_region      36354..37234
                    /rpt_family="L1"
repeat_region      37282..37409
                    /rpt_family="MIR"
repeat_region      38220..38518
                    /rpt_family="Alu"
repeat_region      39945..40266
                    /rpt_family="L2"
repeat_region      41772..42169
                    /rpt_family="MALR"
repeat_region      42170..42339
                    /rpt_family="MIR"
repeat_region      42978..43147
                    /rpt_family="MIR"

repeat_region      43148..43639
                    /rpt_family="MIR"
repeat_region      44708..45006
                    /rpt_family="Retroviral"
repeat_region      45940..46256
                    /rpt_family="Alu"
repeat_region      46330..46564
                    /rpt_family="7SLRNA"
repeat_region      48279..48540
                    /rpt_family="L2"
repeat_region      48927..49030
                    /rpt_family="Alu"
repeat_region      49237..49527
                    /rpt_family="L1"
repeat_region      49528..49613
                    /rpt_family="Alu"
repeat_region      50590..50877
                    /rpt_family="(TAGA)n"
repeat_region      50890..51025
                    /rpt_family="Alu"
repeat_region      53705..53841
                    /rpt_family="MIR"
repeat_region      54158..54335
                    /rpt_family="MIR"
repeat_region      54676..54750
                    /rpt_family="MIR"
repeat_region      57756..57891
                    /rpt_family="L2"
repeat_region      62590..62636
                    /rpt_family="MIR"
repeat_region      63046..63203
                    /rpt_family="L2"
repeat_region      63324..64041
                    /rpt_family="L1"
repeat_region      64113..64173
                    /rpt_family="L2"
repeat_region      64181..64443
                    /rpt_family="L2"
repeat_region      64506..65064
                    /rpt_family="Alu"
repeat_region      66287..66581
                    /rpt_family="L2"
repeat_region      67183..67278
                    /rpt_family="Alu"
repeat_region      67515..67605
                    /rpt_family="MIR"
repeat_region      72685..72966
                    /rpt_family="MIR"
                    /rpt_family="Alu"

Query Match      57.5%; Score 23; DB 11; Length 161652;
Best Local Similarity 74.4%; Pred. No. 91;
Matches 29; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY      1 TCGTGAATTCATGCTGCTATTTCATGAACAATG 39
Db      91650 TGGTTAAATGTAATGTAATGAATTAATGAGACAAATG 91612

RESULT 6
LOCUS      HUAC002565 103911 bp      DNA      PRI      17-MAR-1998
DEFINITION Human Chromosome 16 BAC clone CIT987SK-A-598D4, complete sequence.
ACCESSION      AC002565
NID      92896804
VERSION      AC002565.1 GI:2896804
KEYWORDS      HTG.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 103911)
              Adams,M.D., Loftus,B.J., Zhou,L., Crosby,M., Fuhrmann,J.,
              .AUTHORS

```


CDS
join(825..1085,1254..1745,1901..2397,2519..2743,
2884..3112)
/codon_start=1
/evidence=not_experimental
/product="F20N2.1"
/protein_id="AAC83052.1"
/db_xref="PID:g3953456"
/db_xref="GI:3953456"
/translation="MEESRDILMTLILIESGVSPGDTVSSETPPALVSICAOALLN
LIDPSASEDELPSLPERFICDIAHSVKNLGYINDMSYKKNVLFKTLISGLGFS
YLCEHAETFOCMCLROFLHSEDDSYRLVFLVERLSEISGKRTLAGIARPKM
ETFRDISDDMVAVNDEKDEPDMHIOKVEAVLKLMTSEKSSDSLAKNTANAVDS
SOKDDPVTVSRDLSLSDSCRENSYEDPEFTNVEYELONOHDLLETESGSSO
LCSLESELLELOMAERLDDKRGGSYLBNOQLVYKRONIDLKORTQYITIF
KLKASHSVNTSTSGHDGRVMTITTYIMSSFDDVRLLETKLILLDQLHAVEEP
AKERFHLRTLELDQSLSEIORREDCRLYNELRPPAPARSKYIMIKKITEIN
SRKIDTIOIRISGETRELOLEKNSIOERLSRYAVVDEMTEVVKDPAQVYKLLT
SISIFEQISEKIMLTDRFRETVDYENKLGSLIARQMSLEKILQADDAIKENESLK
K"
CDS
complement(join(3612..3905,3995..4372,4466..4715,
4916..5204,5303..5406,5505..5625,5711..5786,5878..5927,
6459..6633))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F20N2.3"
/protein_id="AAC83053.1"
/db_xref="PID:g3953457"
/db_xref="GI:3953457"
/translation="MSSSDGSSSVSPYHLKTKLOHIREYTKGQSDIEDIFAYKDS
FNDPHTAOKLFLDTPHEVRSKREKKEASPIYVQPSGRGRNPASNSIQGSS
GRNASFRKENGANHTVGRTOAPATKASATVPNTKVSIPASIPSEVSHNAQDD
PSLISARCSKSDQALEITASKQGNQSLPKPDVSESHVTFPHLQVAKGLQNGL
TFGEFDNPFKSVSSNGASGGYSDNESHGTDDESESPITNGITGVASAEES
TVSEDKDYGINSATGAEPYVSHDIYVPEVEKREKALSTETHOIAEGEALSYF
GLVPSLAIQGPVTEAEHQPGNSNPISLIVYPOSSIAATOOTENTLROVYRP
NFPYGYSPYMPPIYHOELSPNGIPCOOSYFGALATPAPAKPYDNTENPTTN
PLYHTSPVASSITSTTTLNSIHSEKASHTESAAVIGQFGFNLQVNPMTIAYQG
QPLGFVYQAGHGMHOPDMAASTYQTLPPPHTTMGEPIGHPIAYQO
PQALITMVNNY"
CDS
complement(join(7027..7176,7296..7378,7472..7704,
7774..7834,7942..8002,8076..8159,8252..8446,8523..8602,
8649..8832,9045..9163,9276..9345,9447..9538,9604..9757))
/note="unknown: similar to uridine kinase (Z99111);
similar to uracil phosphoribosyl transferase (A1022223);
similar to ESTs gb|746383, gb|N99614, and gb|AA979998"
/codon_start=1
/evidence=not_experimental
/product="F20N2.3"
/protein_id="AAC83054.1"
/db_xref="PID:g3953458"
/db_xref="GI:3953458"
/translation="MASKSDVNIETSSKVFSGFHQMDGLASNRPEOMAEHEHGOP
FVIESRGVAGGASGKTVCMDIMQOLHDQRAVVNODSPYHNVNEIYRVADYNP
DHPADPTEOLISMERLKGQAVDINYDKSKNNTPRRYRNPSSVITLLEILIF
HDPYRPLAMNKIFVDAGLSHTKPVNTYVAVAKMRCOTICHEDDVAIDLARIRK
DYEKGRDIATVLDQYSKVPKPAFEDILPKRYADIIIPGSGNHAVIDLVQIHIT
KLQGHDLCKITPNLYVOSTFOIRGMHTLIDSSTKHDFEYSDRLIRLVNGLGHL
LPFERQVVTPTGSVYGVDCKIKCQSVSRSEGENALPACGKIKIKILHRE
GDNGOQCVSLITSPNYLITNGTQLIYERLPSDSEKRVLLDPIIGTSASVO
AIRLISKVPSNIIIFNLNISAPGVAVYCKPRIKITYISELGLNDFRVPYPCG
GEGDKRFGDDP"
CDS
11592..12074
/note="unknown: similar to A1022223: similar to EST
gb|N37483"
/codon_start=1
/evidence=not_experimental
/product="F20N2.4"
/protein_id="AAC83055.1"
/db_xref="PID:g3953459"
/db_xref="GI:3953459"
/translation="MFSSTIRLIVSGFHRQPLKSPVNSPVSFIVSPKFNSESRSNG
TGSRSVAMSSVEKTSQSGALENRASNREKLEQLPVELVIEDVSYQHAGHGKMG

CDS
join(12795..12975,13124..13403,13459..14176)
/note="hypothetical protein: similar to A1022223"
/codon_start=1
/evidence=not_experimental
/product="F20N2.5"
/protein_id="AAC83056.1"
/db_xref="PID:g3953460"
/db_xref="GI:3953460"
/translation="MRELSELNRDQDEEYVWVKYSSKHQILLVGEDEFSCLAT
CFGASNIVASSLSDYKRYVNGCCSKRLDPLSCMSFMTKPDVYRKYNASNL
TUKLGFLLHGADATLHRRPDYRPRPVYITNPFHTGFRKRESPCOIQPAAATL
KNLFKDFLHGASHLRADGEVYSHKNAKPCYINLELASRCLVLIOLEAFERNY
PGYENKKGDSRCDQPLLGECSTFKFRFRAKVELYAEKVEYKSCPRBLSTRAAN
KHAERFESSIHLEPCTRFKFSVLEHNGELQVRYIRTFPHHTYKSSSTRLL
CODVAVQASQEPFERBLRRRGVSHGHNGRVRKMLRNSGSRKLRMRORNLKIR
"
CDS
join(14961..15147,15440..15638,15767..16253,16466..16690,
16860..16958,17122..17236)
/note="similar to A1022223 and AF067215"
/codon_start=1
/evidence=not_experimental
/product="F20N2.6"
/protein_id="AAC83057.1"
/db_xref="PID:g3953461"
/db_xref="GI:3953461"
/translation="MAMRELSQLNRSDQDEEYVWVKYSSKHQILLVGEDEFSHSL
ATLGASNIVASSLSDYKRYVNGCCSKRLDPLSCMSFMTKPDVYRKYNASNL
YRRPDYRPRPVYITNPFHTGFRKRESPCOIQPAAATLKNLFKDFLHGASHL
RADGEVYSHKNAKPCYINLELASRCLVLIOLEAFERNYPGYENKKGDSRCDQPL
LGECSTFKFRFRAKVELYAEKVEYKSCPRBLSTRAANKHAERFESSIHLEPCTRFKFSV
LEHNGELQVRYIRTFPHHTYKSSSTRLLCODVAVQASQEPFERBLRRRGVSHGHNGRVRKMLRNSGSRKLRMRORNLKIR
"
CDS
join(17736..17910,18005..18229,19477..19991)
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F20N2.7"
/protein_id="AAC83058.1"
/db_xref="PID:g3953462"
/db_xref="GI:3953462"
/translation="WOLFSPSFLFKNLVSVKLVSCRTNLTYSFQORIVLKMDSDE
KSNKRAKIASAKPGVSVSIQGNDDLVLEGEIDALEITRELRKRVCMITTITVYA
APQOPQPHMGQYQMPPARCTCIPNAGVLFCLIVISVTCNANSGMISDLC
KHSDDPNLCSITSRPESGEFAGTSIOETIATISANASATRSYTIKOKLSNDE
PAIEDTLEDOCKYQDAVBOUDSISMADAHQDVYVWLSAATSAIESGSGALGSA
GNDELQSRNEVFLCKNALMINKMLT"
CDS
join(22523..22641,22783..22931,23048..23166,23311..23913)
/note="similar to MEL-26 (U67737)"
/codon_start=1
/evidence=not_experimental
/product="F20N2.8"
/protein_id="AAC83059.1"
/db_xref="PID:g3953463"
/db_xref="GI:3953463"
/translation="MTDSAYRVDTISRLAQMRHINLSSSTYRKSDPFKMGLANHLSY
EKSMLNVLKLYPNTLRENPPVAFALRYVSEKRALSHPEYIDIRIKNEP
IMTEVLEGLKTIIDVDFDLKVISQDSGELYSWANGSENQOQVAVYSLGMLTE
SITDITINSDGSIQAHRAVLAKRSYVFSMFLHDKRELSTINVLQNLQAF
LSYVGINQWEDFLIHLALQAAEKYDIALDKACHLSKLDLIDITNVLERLQNAVL
YOLPELASQKRYLVKFGKIFEIDEFENIMQCDRLISEIFHEVISTWKGF"
CDS
complement(join(24535..24616,24851..24962,25135..25159))
/note="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="F20N2.9"
/protein_id="AAC83060.1"
/db_xref="PID:g3953464"
/db_xref="GI:3953464"
/translation="MSTAPEEASAEBOYSLSSKLEEDMKRISDSKKKKAMVYVIGSC
RSPFALRISVLESDVSTEDLKSILM"


```
repeat_region 56261..56547
                /rpt_family="ALU"
misc_feature 57863..58519
              /gene="WUGSC:H.263A.1"
              /note="match to human 5' EST T77253 (NID:9694456) and 3'
              EST T77470 (NID:9694673), reads from opposite ends of the
              same clone"
exon complement(58576..58620)
          /gene="WUGSC:H.263A.1"
          /note="Grail prediction, score = 96"
          /evidence=not_experimental
repeat_region 59652..59931
              /rpt_family="ALU"
              /evidence=not_experimental
repeat_region 63783..64081
              /rpt_family="ALU"
              /evidence=not_experimental
exon 69913..70050
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 88"
      /evidence=not_experimental
repeat_region 74739..75028
              /rpt_family="ALU"
              /evidence=not_experimental
repeat_region 77367..77447
              /rpt_family="ALU"
              /evidence=not_experimental
exon 80069..80361
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 94"
      /evidence=not_experimental
exon 84245..84412
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 80"
      /evidence=not_experimental
repeat_region 82912
              /rpt_family="ALU"
              /evidence=not_experimental
repeat_region 83688
              /rpt_family="ALU"
              /evidence=not_experimental
exon 99855..99856
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 96"
      /evidence=not_experimental
repeat_region 98415
              /rpt_family="ALU"
              /evidence=not_experimental
repeat_region 99192
              /rpt_family="ALU"
              /evidence=not_experimental
exon 100410..100538
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 82"
      /evidence=not_experimental
exon 105514..105666
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 98"
      /evidence=not_experimental
exon 108897..109182
      /gene="WUGSC:H.263A.1"
      /note="Grail prediction, score = 83"
      /evidence=not_experimental
repeat_region 108897..109182
              /rpt_family="ALU"
              /evidence=not_experimental
BASE COUNT 33497 a 20778 c 21335 g 35248 t
ORIGIN
Query Match 56.0%; Score 22.4; DB 10; Length 110858;
Best Local Similarity 72.5%; Pred. No. 1.5e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1 TGCCTGAATTCATGCTGATTTCAAGAACAAATCA 40
    ||| ||||| ||||| ||| ||| ||| ||||| |||
```

```
DB 27259 TGCAGAAATTTTATGCGGCTTTTATACACAAATTA 27298
RESULT 10
AC004127
LOCUS
DEFINITION Homo sapiens chromosome 11 clone pD741n15, WORKING DRAFT SEQUENCE.
ACCESSION AC004127
VERSION 92842777
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 170215)
            Evans,G.A., Athanasiou,M., Basil,M., Bradbury,P., Brignac,S.,
            Buneslter,R., Davis,C., English,C., Franklin,T.L., Garner,H.R.,
            Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
            Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P.,
            Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
            Ward,T. and Wilson,R.
            template
TITLE      Unpublished
JOURNAL   2 (bases 1 to 170215)
REFERENCE Evans,G.A., Athanasiou,M., Basil,M., Bradbury,P., Brignac,S.,
AUTHORS   Buneslter,R., Davis,C., English,C., Franklin,T.L., Garner,H.R.,
            Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
            Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P.,
            Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D.,
            Ward,T. and Wilson,R.
            Direct Submission
TITLE      Submitted (06-FEB-1998) Genome Science & Technology Center,
JOURNAL   University of Texas Southwestern Medical Center, 5323 Harry Hines
            Blvd, Dallas, TX 75235-8591, USA
COMMENT   * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 2519: contig of 2519 bp in length
            * 2 4982: contig of 2463 bp in length
            * 3 4983 13767: contig of 8785 bp in length
            * 4 13768 32352: contig of 18585 bp in length
            * 5 32353 55696: contig of 23344 bp in length
            * 6 55697 86119: contig of 30423 bp in length
            * 7 86120 170215: contig of 84096 bp in length.
FEATURES
    source
        1..170215
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="pD741n15"
            /chromosome="11"
BASE COUNT 47532 a 37634 c 37608 g 47376 t 65 others
ORIGIN
Query Match 56.0%; Score 22.4; DB 34; Length 170215;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 1 TGCCTGAATTCATGCTGATTTCAAGAACAAATCA 40
    ||| ||||| ||| ||| ||||| ||| ||| ||| ||||| |||
Db 78624 TATGAAATGTTAGTGCAGATATTGAAGACCAATCA 78663
RESULT 11
AC006740/c AC006740 200965 bp DNA HTG 25-FEB-1999
LOCUS
```

```

DEFINITION  Caenorhabditis elegans clone Y38B5, WORKING DRAFT SEQUENCE, 12
unordered pieces.
ACCESSION   AC006740
NID         94309800
VERSION     AC006740.2  GI:4309800
KEYWORDS    HTG: HTGS_PHASE1.
SOURCE      Caenorhabditis elegans.
ORGANISM    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida;
            Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 200965)
AUTHORS     Waterston,R.H.
TITLE       The sequence of Caenorhabditis elegans clone
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 200965)
AUTHORS     Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (23-FEB-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Mar 1, 1999 this sequence version replaced g1:4263198.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 33116: contig of 33116 bp in length
            * 33117 33125: gap of unknown length
            * 33126 36942: contig of 3817 bp in length
            * 36943 36951: gap of unknown length
            * 36952 39003: contig of 2052 bp in length
            * 39004 39012: gap of unknown length
            * 39013 46010: contig of 6998 bp in length
            * 46011 46019: gap of unknown length
            * 46020 59290: contig of 13271 bp in length
            * 59291 59299: gap of unknown length
            * 59300 130771: contig of 71472 bp in length
            * 130772 130780: gap of unknown length
            * 130781 155202: contig of 24422 bp in length
            * 155203 155211: gap of unknown length
            * 155212 160675: contig of 5435 bp in length
            * 160676 170159: contig of 9484 bp in length
            * 170160 170168: gap of unknown length
            * 170169 173251: contig of 3083 bp in length
            * 173252 173260: gap of unknown length
            * 173261 176650: contig of 3390 bp in length
            * 176651 176658: gap of unknown length
            * 176659 200965: contig of 24307 bp in length.
            *
            * Location/Qualifiers
            * 1..200965
              /organism="Caenorhabditis elegans"
              /db_xref="taxon:6239"
              /clone="Y38B5"
BASE COUNT  61403 a 37295 c 36272 g 64715 t 1280 others
ORIGIN
Query Match 56.0%; Score 22.4; DB 34; Length 200965;
Best Local Similarity 72.5%; Pred. No. 1.4e+02;
Matches 29; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 TGCTGTAATCTAATGCTGATTCATGACGAACAATGA 40
   ||||| ||||| ||||| ||||| ||||| |||||
Db 109845 TTCTGAATAATTTCATGATTCATGACGACATGA 109806

RESULT  12
CELK08B12/c
LOCUS   CELK08B12 37678 bp DNA INV 16-APR-1997

```

```

DEFINITION  Caenorhabditis elegans cosmid K08B12.
unordered pieces.
ACCESSION   U97001
NID         91938418
VERSION     U97001.1  GI:1938418
KEYWORDS    HTG: HTGS_PHASE1.
SOURCE      Caenorhabditis elegans strain-Bristol N2.
ORGANISM    Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida;
            Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 37678)
AUTHORS     Wilson,R., Almscough,R., Anderson,K., Baynes,C., Berts,M.,
            Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J.,
            Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
            Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
            Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Lalster,N.,
            Latreille,P., Lightning,J., Lloyd,C., McKurray,A., Mortimore,B.,
            O'Callaghan,M., Parsons,J., Percy,C., Rifkin,L., Roopra,A.,
            Saunders,D., Showkeen,R., Smaildon,N., Smith,A., Sonhammer,E.,
            Staden,R., Sulston,J., Thierly-Mieg,U., Thomas,K., Vaudin,M.,
            Vaughan,K., Waterston,R., Watson,A., Wellstock,L.,
            Wilkinson-Spratt,J. and Wohlmann,P.
TITLE       2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
JOURNAL     Nature 368 (6466), 32-38 (1994)
MEDLINE    94150718
REFERENCE   2 (bases 1 to 37678)
AUTHORS     Becker,M. and Wohlmann,P.
TITLE       The sequence of C. elegans cosmid K08B12
JOURNAL     Unpublished (1997)
REFERENCE   3 (bases 1 to 37678)
AUTHORS     Waterston,R.
TITLE       Direct Submission
JOURNAL     Submitted (10-APR-1997)
COMMENT     Submitted by:
            Genome Sequencing Center
            Department of Genetics, Washington University,
            St. Louis, MO 63110, USA, and
            Sanger Centre, Hinxton Hall
            Cambridge CB10 1RQ, England
            e-mail: iw@nematode.wustl.edu and jess@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' end of this cosmid lies in gap:3' cosmid is F31F7, 7400 bp
overlap. Actual start of this cosmid is at base position 1 of
CELK08B12; actual end is at 33279 of CELK08B12

NOTES:
Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1..37678
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="V"
/clone="K08B12"
5609. .8033
gene

```



```

source
1. 236496
   /organism="Caenorhabditis elegans"
   /db_xref="taxon:6239"
   /clone="Y49F68"

BASE COUNT      73408 a 44340 c 44572 g 74176 t
ORIGIN

Query Match      55.5%; Score 22.2; DB 34; Length 236496;
Best Local Similarity 77.1%; Pred. No. 1.5e+02;
Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY      6 TAAATCTAATGCTGATATTCATGAACAATGA 40
Db      54690 TAAATTCAAATCTAATATTTAAACAATGA 54656

RESULT 14
LOCUS      CELY49F6C 19566 bp DNA INV 06-FEB-1999
DEFINITION Caenorhabditis elegans cosmid Y49F6C.
ACCESSION AF125463
VERSION    AF125463.1 GI:4226153
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Wilson, R., Alnscough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Lalister, N.,
Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Riker, L., Roopra, A.,
Saunders, D., Showkhen, R., Smaldon, N., Smith, A., Sonhammer, E.,
Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Sproat, J. and Wohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
MEDLINE
94150718
REFERENCE
2 (bases 1 to 19566)
AUTHORS
Miller, N. and Le, T.
TITLE
The sequence of C. elegans cosmid Y49F6C
JOURNAL
REFERENCE
Unpublished (1998)
AUTHORS
3 (bases 1 to 19566)
TITLE
Waterston, R.
JOURNAL
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
e-mail: rwenem@wustl.edu and jesse@sanger.ac.uk

```

FEATURES

source

This segment of YAC Y49F6 was sequenced to span the gap between F14D2 and T06D4. The 5' clone is F14D2, 200 bp overlap; 3' clone is T06D4, 200 bp overlap. Actual start of YAC Y49F6 is at base position 16839 of CELY49F6; actual end is at 14918 of CELY49F6.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

NEIGHBORING COSMID INFORMATION:

```

1. 19566
   /organism="Caenorhabditis elegans"
   /strain="Bristol N2"
   /db_xref="taxon:6239"
   /clone="Y49F6C"
   /chromosome="II"
   /complement(1360..3975)
   /gene="Y49F6C.6"
   /complement(join(1360..1499,1782..1985,2081..2179,
3387..3715,3863..3975))
   /gene="Y49F6C.6"
   /codon_start=1
   /evidence=not-experimental
   /protein_id="AAD12861.1"
   /db_xref="PID:94226155"
   /db_xref="GI:4226155"
   /translation="MGLTAMGLTWTGVLATMGATTKISQHFKEFRNSQKVL
EHRFPFPISLKRESQEPNPSDPIVLQYSTTVTPPTPNNSKPSISQKX
TQFLNPSNPVLLQYLSTTVTPPTPNNSKPSISQKPTPLSPRIYLKHEH
CRGALSDVNGAAGAVCAKLVKTRHYLRKISANKRLFPLECYTHYTKNOY
SPERNANOKGKRMPEMVGALLPELNNVIERISCHSSSSFSARLSAIEYEMAR
ALQILHNGVETIDREGG"
   /complement(5272..6211)
   /gene="Y49F6C.7"
   /complement(join(5272..5756,5891..5939,5988..6104,
6137..6211))
   /gene="Y49F6C.7"
   /codon_start=1
   /evidence=not-experimental
   /protein_id="AAD12867.1"
   /db_xref="PID:94226161"
   /db_xref="GI:4226161"
   /translation="MPPEEHQNLSENYQAPARTVIFIPRNELYAOIOVQDINROEI
RAREEDQRLRLDREMDNRIRIEMNIRAVLENDQSGSFLFLHOFSLISLNP
KSRPFRTERPKRPKRPKRPKRPKRPKRPKRPKRPKRPKRPKRPKRPKRPK
YRPGRVHRAQTTHALIFESVNDRLRVHGAETLEPVHRSASFENLFFEQLL
YAFLOFSDNGSRAPSRNRNGAG"
   /complement(7296..8106)
   /gene="Y49F6C.8"
   /complement(join(7296..7577,7848..7979,8029..8106))
   /gene="Y49F6C.8"
   /codon_start=1
   /evidence=not-experimental
   /protein_id="AAD12860.1"
   /db_xref="PID:94226154"
   /db_xref="GI:4226154"
   /translation="MAQNPNSFTDSVGMKINLVSIRAGFLFELTIKQPYQNSLQRI
QEEHREKQYLROALIDLADESEEDGQPLDEYTSKMAELTIEENNRKSFNIV
NEORTGSKDRQHMEEIRIQMWTKEDEBELSGLAVMKOTIOESWKDAADCEKRSK
NNQ"
   /gene="Y49F6C.5"
   /join(9324..9725,9785..9932,9978..10159)
   /gene="Y49F6C.5"
   /note="contains similarity to BTB (also known as BR-C/TLK)
domains (Pfam:PF00651, Score=58.8, E=1.2e-13, N=1)"
   /codon_start=1
   /evidence=not-experimental

```

gene

CDS

[illegible]

Job time: 4510 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:21:51 ; Search time 133.08 Seconds

(without alignments)
75.201 Million cell updates/sec

Title: US-09-325-095-2122.SEQ

Perfect score: 40

Sequence: 1 TCGTGAATTCATGCTGATTCATGAACAATGA 40

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	22	55.0	3498	1	Q57012	Ptdins 3-kinase 11
2	22	55.0	1440	1	V74847	Staphylococcus aur
3	21.6	54.0	110000	1	V21209_11	Continuation (12 o
4	21.6	54.0	110000	1	V21209_12	Continuation (13 o
5	21.4	53.5	652	1	V74927	Staphylococcus aur
6	21.2	53.0	1192	1	T08635	HPV E6-binding pro
7	21	52.5	110000	1	T58840_0	Mycoplasma genitali
8	21	52.5	184	1	V11694	Divericin A DNA.
9	21	52.5	1431	1	V53139	Helicobacter seryl
10	21	52.0	593	1	X17718	Divericin A signa
11	20.8	52.0	2859	1	T08960	Growth factor-acti
12	20.8	52.0	110000	1	T58840_1	Human 26S proteaso
13	20.8	52.0	786	1	V28681	Continuation (2 of
14	20.6	51.5	4810	1	V70887	Ripening banana pu
15	20.6	51.0	1590	1	V27169	Plasmid pTK-fus-1a
16	20.4	50.5	288	1	V88567	Bacillus thuringie
17	20.2	50.0	3412	1	Q51156	EST clone FK295. N
18	20	50.0	3207	1	Q51155	Human p110 CDNA. R
19	20	50.0	20	1	Q59020	p110 CDNA. Recombi
20	20	50.0	20	1	Q59020	Antisense PCR prim
21	20	50.0	20	1	Q59021	Sense PCR primer f
22	20	50.0	3498	1	Q57012	Ptdins 3-kinase 11
23	20	50.0	2766	1	T75488	DNA for Bacillus s
24	20	50.0	192	1	V88282	EST clone GE410. N
25	20	50.0	3501	1	X13430	Enterococcus faeca
26	20	50.0	9767	1	X13067	Enterococcus faeca
27	20	50.0	725	1	X13997	H. pylori GHP 876
28	20	50.0	110000	1	X20248_06	Continuation (7 of
29	20	50.0	111309	1	X20250	Borrelia burgdorfe
30	19.8	49.5	3932	1	Q64112	cytR5 gene. Isola
31	19.8	49.5	2288	1	Q83526	P. falciparum EBL-
32	19.8	49.5	4948	1	T42134	12C-1 gene encodin
33	19.8	49.5	3934	1	T68434	H. pylori secreted
34	19.8	49.5	873	1	T68271	Plasmid encoding ebl-1 g
35	19.8	49.5	2288	1	T72894	DNA encoding Bacil
36	19.8	49.5	3934	1	T95051	MSRV-1 clone LTRGA
37	19.8	49.5	2503	1	T96475	Clone LTRGAG12 DNA
38	19.8	49.5	2503	1	V43169	DNA encoding a Cry
39	19.8	49.5	3934	1	V83927	H. pylori secreted
40	19.8	49.5	870	1	X30444	Type 5 17-beta-hyd
41	19.6	49.0	700	1	T63475	Novel protein from
42	19.6	49.0	2095	1	V12251	Enterococcus faeca
43	19.6	49.0	17764	1	X13238	

44 19.6 49.0 11879 1 X13118
45 19.4 48.5 1262 1 Q63213

Enterococcus faeca
Transcription acti

ALIGNMENTS

```

RESULT 1
ID 057012/c
AC 057012:
DE 31-AUG-1994 (first entry)
KW Ptdins 3-kinase 110 kD catalytic subunit cDNA.
KW 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;
KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;
KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
KW blood vessel plaques; ss.
OS Bos taurus.
FH Key
FT Location/Qualifiers
FT cds
FT 1..3207
FT /tag= A
FT /product= p110
PD MO9403609-A.
PD 17-FEB-1994.
PD 05-AUG-1993; G01651.
PD 05-AUG-1992; GB-016654.
PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
DR WPI: 94-065697/08.
DR P-PSDB: R46294.
DR Eukaryotic cells transformed with mammalian phospholipid or
PT protein kinase DNA - useful in assays for compounds involved in
PT cell growth regulation and for treating cancers
PS Disclosure: Fig 1; 71pp: English.
CC This sequence encodes the 110 kD catalytic subunit of the phosphatidyl
CC inositol (Ptdins) 3-kinase. This sequence was transformed into
CC Schizosaccharomyces pombe cells under the regulatory control of the
CC nmt promoter in an embodiment of the invention. In the presence of
CC thiamine the promoter is inactive and the cells carrying the Ptdins
CC catalytic subunit plasmid grow as the parental strain. In the absence
CC of thiamine the nmt promoter functions and the Ptdins 3-kinase
CC catalytic subunit is induced. Ptdins activity is substantially
CC increased under these conditions. Cells containing constructs such as
CC this, are useful in assays for detecting compounds involved in cell
CC growth regulation. It is also used as the basis for detecting
CC compounds for treating cancers and the formation of blood vessel
CC plaques.
SQ Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 55.0%; Score 22; DB 1; Length 3498;
Best Local Similarity 73.7%; Pred. No. 17;
Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 TCGTGAATTCATGCTGATTCATGAACAAT 38
DB 3325 TCGTGAATTCATGCTGATTCATGAACAAT 3288

RESULT 2
ID V74847
AC V74847:
DE 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #536.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key
FT Location/Qualifiers
FT misc_feature 121..180
FT /tag= a

```

/note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP-786519-A2.

30-JUL-1997.

PF 07-JAN-1997; 100117.

PR 05-JAN-1996; US-009861.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA;

DR MPI: 97-374922/35.

Polynucleotide(s) and proteins derived from *Staphylococcus aureus* - stored on computer readable medium and used in the production of anti-S.aureus vaccines

PS Claim 1, Page 1470-1471; 3271pp; English.

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used

CC for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S.aureus DNA sequences contained on the computer readable medium.

CC Sequence 1440 BP: 393 A: 286 C: 211 G: 507 T:

SQ

Query Match 55.0%; Score 22; DB 1; Length 1440;
Best Local Similarity 70.0%; Pred. No. 15;
Matches 28; Conservative 1; Mismatches 11; Indels 0; Gaps 0;

OY 1 TGCTGAATTCATGCTGATTCATGAACAATGA 40

Db 58 TTCTCTATTTCTACGCTTGAMATTGCATCAACAATCA 97

RESULT 3

Continuation (12 of 17) of V21209 from base 1100001 (Methanococcus jannaschii circular

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP Fragment Name Begin End

WP V21209_00	1	110000
WP V21209_01	100001	210000
WP V21209_02	200001	310000
WP V21209_03	300001	410000
WP V21209_04	400001	510000
WP V21209_05	500001	610000
WP V21209_06	600001	710000
WP V21209_07	700001	810000
WP V21209_08	800001	910000
WP V21209_09	900001	1010000
WP V21209_10	1000001	1110000
WP V21209_11	1100001	1210000
WP V21209_12	1200001	1310000
WP V21209_13	1300001	1410000
WP V21209_14	1400001	1510000
WP V21209_15	1500001	1610000
WP V21209_16	1600001	1664976

Query Match 54.0%; Score 21.6; DB 1; Length 110000;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 3 CTGTAATTCATGCTGATTCATGAACAAT 38
Db 108749 CTTTAATTCATGCTATTAATGACGCTAACAAT 108714

RESULT 4

Continuation (13 of 17) of V21209 from base 1200001 (Methanococcus jannaschii circular

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP Fragment Name Begin End

WP V21209_00	1	110000
WP V21209_01	100001	210000
WP V21209_02	200001	310000
WP V21209_03	300001	410000
WP V21209_04	400001	510000
WP V21209_05	500001	610000
WP V21209_06	600001	710000
WP V21209_07	700001	810000
WP V21209_08	800001	910000
WP V21209_09	900001	1010000
WP V21209_10	1000001	1110000
WP V21209_11	1100001	1210000
WP V21209_12	1200001	1310000
WP V21209_13	1300001	1410000
WP V21209_14	1400001	1510000
WP V21209_15	1500001	1610000
WP V21209_16	1600001	1664976

Query Match 54.0%; Score 21.6; DB 1; Length 110000;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 3 CTGTAATTCATGCTGATTCATGAACAAT 38

Db 8749 CTTTAATTCATGCTATTAATGACGCTAACAAT 8714

RESULT 5

ID V74927/c

AC V74927: standard; DNA; 652 BP.

DT 16-MAR-1999 (first entry)

DE *Staphylococcus aureus* contig SEQ ID #616.KW Computer readable medium; vaccine; *S.aureus* infection; immunodetection;

KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;

KW toxic shock syndrome; ds.

OS *Staphylococcus aureus*.

EP-786519-A2.

PN 30-JUL-1997.

PF 07-JAN-1997; 100117.

PR 05-JAN-1996; US-009861.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA;

DR MPI: 97-374922/35.

Polynucleotide(s) and proteins derived from *Staphylococcus aureus* - stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

PS Claim 1, Page 1545; 3271pp; English.

CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,

CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock,
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S. aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 652 BP; 197 A; 124 C; 76 G; 255 T;

Query Match 53.5%; Score 21.4; DB 1; Length 652;
 Best Local Similarity 71.8%; Pred. No. 23;
 Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 2 GCTGAATTTCTAATGCTGTAATTCATGAACAATGA 40
 DB 215 GCTGAAGTTAATGAAGAAATTCATACCAACTGA 177

RESULT 6
 T08635
 ID T08635 standard; CDNA: 1192 BP.
 AC T08635;
 DT 08-MAY-1996 (first entry)
 DE HPV E6-binding protein CDNA clone SD-7.
 KW E6 binding protein; E6-BP; human papilloma virus; HPV; antagonist;
 KW immunogen; Hela; cervical cancer; transgenic animal; gene therapy;
 KW diagnosis; ss.
 CS Homo sapiens.
 FH Key
 FT cds Location/Qualifiers
 FT 173..832
 FT /*tag= a
 FT MO9602000-A1.
 PN 25-JAN-1996.
 PD 15-JUN-1995; U07780.
 PR 08-JUL-1994; US-273059.
 PA (NEMO-) NEW ENGLAND MED CENT HOSPITALS.
 PI Androphy E, Chen JJ;
 DR WPI: 96-097705/10.
 DR P-PEDB; R77660.
 PT Papilloma virus E6-binding proteins - used to develop prods. for the
 PT study, diagnosis and treatment of papilloma virus infection
 PS Claim 11; Page 53; 71pp; English.
 CC A CDNA clone (T08635), designated SD-7, codes for the N-terminal
 CC portion (R77660) of E6-BP-SD-7, a protein that specifically binds to
 CC human papilloma virus (HPV) protein E6. SD-7 and other E6-BP-encoding
 CC clones (see T08636-41) were identified using a two-hybrid assay in
 CC which yeast expressing the HPV-16 E6 gene fused to the BPV E2 DNA-
 CC binding domain was transformed with a vector carrying Hela CDNA and a
 CC VP16 transcription activation domain. These novel CDNA clones
 CC (deposited as a library of PRS306 plasmids as ATCC 75827) are used to
 CC produce recombinant E6-BPs in host cells, as primers or probes e.g. to
 CC detect genetic lesions, to raise transgenic animals, and in gene
 CC therapy.
 SQ Sequence 1192 BP; 382 A; 191 C; 260 G; 359 T;

Query Match 53.0%; Score 21.2; DB 1; Length 1192;
 Best Local Similarity 76.5%; Pred. No. 29;
 Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 4 TGTAAATTTCTAATGCTGTAATTCATGAACAATA 37
 DB 959 TGTAGATTAATTTGCTTTTACGAAAAAAA 992

RESULT 7
 T58840_0/c
 WP Sequence split into 6 fragments LOCUS T58840 Accession T58840
 WP Fragment Name Begin End
 WP T58840_0 1 110000
 WP T58840_1 100001 210000
 WP T58840_2 200001 310000

WP T58840_3 300001 410000
 WP T58840_4 400001 510000
 WP T58840_5 500001 580073
 ID T58840 standard; DNA: 580073 BP.
 AC T58840;
 DT 27-MAR-1997 (first entry)
 DE Mycoplasma genitalium genome.
 KW M. genitalium; DNA; DNA gyrase; origin of replication;
 KW megabase shotgun sequencing method; open reading frame; ORF; ss.
 OS Mycoplasma genitalium.
 FH Key
 FT cds Location/Qualifiers
 FT 8552..9184
 FT /*tag= a
 FT /label= MG006
 FT /note= "Previously identified as MORF-20076, the
 FT encoded protein shows 27.59 percentage
 FT identity to thymidylate kinase (CDC8)
 FT from Saccharomyces cerevisiae"
 FT 11252..12040
 FT /*tag= b
 FT /label= MG009
 FT /note= "Previously identified as MORF-20078, the
 FT encoded protein shows 35.43 percentage
 FT identity to the Bacillus subtilis hypothetical
 FT protein covered in accession number
 FT GB:D26185_102"
 FT 12069..12725
 FT /*tag= c
 FT /label= MG010
 FT /note= "Previously identified as MORF-20079, the
 FT encoded protein shows 25.73 percentage
 FT identity to DNA primase (dnaB) from
 FT Clostridium acetobutylicum"
 FT complement (13570..14247)
 FT /*tag= d
 FT /label= MG012
 FT /note= "Previously identified as MORF-20080, the
 FT encoded protein shows 31.50 percentage
 FT identity to the ribosomal protein S6
 FT modification protein (rlnk) from Escherichia
 FT coli"
 FT complement (14396..15217)
 FT /*tag= e
 FT /label= MG013
 FT /note= "Previously identified as MORF-19823, MORF-20080
 FT and MORF-20081, the encoded protein shows 33.04
 FT percentage identity to 5,10-methylene-tetra-
 FT hydrofolate dehydrogenase (fold) from E. coli"
 FT 17474..19243
 FT /*tag= f
 FT /label= MG015
 FT /note= "Previously identified as MORF-20084, the
 FT encoded protein shows 32.23 percentage
 FT identity to transport ATP-binding protein
 FT (msbA) from E. coli"
 FT 26478..27344
 FT /*tag= g
 FT /label= MG023
 FT /note= "Previously identified as MORF-20092, the
 FT encoded protein shows 45.96 percentage
 FT identity to fructose-bisphosphate aldolase
 FT (tsr) from B. subtilis"
 FT 27345..28448
 FT /*tag= h
 FT /label= MG024
 FT /note= "Previously identified as MORF-19826 and
 FT MORF-20093, the encoded protein shows 46.84
 FT percentage identity to GTP-binding protein
 FT from E. coli"
 FT 36987..38978
 FT /*tag= i
 FT /label= MG032
 FT /note= "Previously identified as MORF-20099, the


```
FT mat_peptide /tag- b
FT 126.184
FT /tag- c
FT 114.144
FT p_limer_bind /tag- d
FT
PD WC0909639-A1.
PD 12-MAR-1998.
PD 05-SEP-1997; U15609.
PR 05-SEP-1996; US-026257.
PA (UNAL-) UNIV ALBERTA.
PI Franz CM, Greer GG, Leisner JT, McCormick JK, McMullen LM,
PI Pknox J, Poono, Stiles ME, Van Belkum MJ, Vederas JC,
PI Morobo RJ, Morobo RW;
DR WPI: 98-193319/17.
DR P-PSDB; W59007.
PT Bacterial growth inhibiting peptide(s) enterocin 900 or brochocin C
PT - used to inhibit growth of susceptible bacteria in e.g. foodstuff,
PT laying animal, food preparation area or fermentation vessel
PS Example 2: Page 115; 174pp; English.
CC This DNA sequence codes for divergicin A (see W59007), a bacteriocin
CC produced by the meat lactic acid bacterium, Carnobacterium divergens.
CC The invention provides a secretion vector in which the structural and
CC immunity genes of other bacteriocins, such as carnobacteriocin B2
CC (see W59029-31), colicin V, leucocin A and brochocin-C (see W59021),
CC are fused to the divergicin A signal peptide (see W59010) gene
CC sequence, allowing production and secretion of the bacteriocin(s) by
CC host and heterologous bacteria. Bacteriocin cassettes encoding the
CC divergicin A signal peptide and two or more bacteriocins are used to
CC tailor the antibacterial spectrum of a producer strain to target a
CC range of spoilage or pathogenic bacteria, e.g. Escherichia coli and
CC Salmonella. Using the technology of multiple bacteriocin production
CC and delivery using lactic acid bacteria, a range of bacteriocins can
CC be produced in situ in human food, animal feed, the gastrointestinal
CC tract of humans and animals, and in the female genital tract.
SQ Sequence 184 BP; 60 A; 24 C; 37 G; 63 T;
```

```
Query Match 52.5%; Score 21; DB 1; Length 184;
Best Local Similarity 73.0%; Pred. No. 29;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
```

```
OY 1 TGCCTAATCTCATGCTGATTTCATGAACAAA 37
| | | | | | | | | | | | | | | | | |
DB 121 TGCCTAATCTCATGCTGATTTCATGAACAAA 157
```

```
RESULT 9
V53139
ID V53139 standard; DNA: 1431 BP.
AC V53139;
DT 12-NOV-1998 (first entry)
DE Helicobacter seryl-tRNA synthetase encoding DNA.
KW Helicobacter: aminoacyl-tRNA synthetase; isoleucyl-tRNA synthetase;
KW methionyl-tRNA synthetase; leucyl-tRNA synthetase; valyl-tRNA synthetase;
KW lysyl-tRNA synthetase; seryl-tRNA synthetase; human gastric mucosa;
KW chronic gastritis; antimicrobial compound; antisense inhibition;
KW H pylori; bacterium; enzyme; ds.
OS Helicobacter pylori.
FH Key Location/Qualifiers
FT CDS 80..1327
FT /tag- a
FT /product= "seryl-tRNA synthetase"
FT
PD US801013-A.
PD 01-SEP-1998.
PR 26-MAY-1995; 451715.
PR 26-MAY-1995; US-451715.
PA (CUBI-) CUBIST PHARM INC.
PI Houman F, Qiu Y, Schimmel PR, Shen X, Tao J;
PI WPI: 98-494769/42.
DR P-PSDB; W75914.
PT DNA encoding Helicobacter amino-acyl-tRNA synthetase proteins -
PT useful for recombinant production of the enzyme, anti-sense
PT constructs and hybridisation probes, and construction of tester
```

```
PT strains to test for inhibitors of the enzyme
PS Claim 5; Columns 95-100; 76pp; English.
CC This DNA encodes a Helicobacter seryl-tRNA synthetase. The invention
CC provides nucleic acid sequences (V53134 to V53139) encoding Helicobacter
CC aminoacyl-tRNA synthetases (hatRNAs) (W75909 to W75914), where the
CC aminoacyl-tRNA synthetase is isoleucyl-tRNA synthetase, methionyl-tRNA
CC synthetase, leucyl-tRNA synthetase, valyl-tRNA synthetase, lysyl-tRNA
CC synthetase or seryl-tRNA synthetase, respectively and at least a portion
CC of the hatRNAs has catalytic activity or binding function. The nucleic
CC acids and recombinant cells containing the nucleic acids can be used for
CC recombinant production of the enzymes. The nucleic acids can also be used
CC as hybridisation probes in experiments to identify hatRNAs sequences.
CC H. pylori is bacteria that infects human gastric mucosa, leading to
CC chronic gastritis. The amino acyl tRNA synthetases are used by the
CC bacterium during protein synthesis, and as such can be a target for
CC antimicrobial compounds. The tester strains can be used to screen for
CC such compounds, and antisense constructs based on the nucleic acids can
CC be used for antisense inhibition of the enzyme. The proteins themselves
CC can be used to raise antibodies, which in turn can be used for
CC purification and study of the enzyme. Screening of inhibitors of the
CC enzyme, as well as antisense antimicrobial activity through the nucleic
CC acids are novel approaches to inhibition activity against the bacteria,
CC as many present antibiotics used have side-effects and the bacteria are
CC becoming resistant to them.
SQ Sequence 1431 BP; 498 A; 271 C; 320 G; 342 T;
```

```
Query Match 52.5%; Score 21; DB 1; Length 1431;
Best Local Similarity 82.8%; Pred. No. 35;
Matches 24; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
```

```
OY 11 TCTAATGCTGATTTCATGAACAAATG 39
| | | | | | | | | | | | | | | |
DB 619 TTTAATGCTGATTTCATGAACAAATG 647
```

```
RESULT 10
X17718
ID X17718 standard; DNA: 184 BP.
AC X17718;
DT 25-MAY-1999 (first entry)
DE Divergicin A signal peptide encoding DNA.
KW Enterocin 900; brochocin-C; bacterial growth; microorganism; inhibit;
KW food; meat; fermentation; bacteriocin; mastitis; secretory vector; lca;
KW leucocin A; lcad; enzyme; probiotic; L. gelidum; Listeria; Lactobacillus;
KW sulphide-producing; spoilage; sensory property; smell; taste; infection;
KW ground beef; silage; Salmonella; poultry; intestine; therapeutic; colour;
KW mucosal tissue; vacuum-packed; storage; divergicin A; ss.
OS Carnobacterium divergens.
FH Key Location/Qualifiers
FT RBS 24..28
FT /tag- a
FT 38..184
FT CDS 38..125
FT /tag- b
FT sig_peptide 38..125
FT /tag- c
FT primer_bind 114..144
FT /tag- d
FT
PD WC0902555-A1.
PD 21-JAN-1999.
PD 09-JUL-1998; U14331.
PR 05-SEP-1997; US-924629.
PR 09-JUL-1997; US-052072.
PA (UNAL-) UNIV ALBERTA.
PI Franz C, Greer GG, Leisner JT, McCormick JK, McMullen LM,
PI Poono A, Stiles ME, Van Belkum MJ, Vederas JC, Morobo RJ,
PI Morobo RW;
DR WPI: 99-120779/10.
DR P-PSDB; W94865.
PT New bacteriocins enterocin 900 and brochocin-C - useful as food
PT preservatives and therapeutic antimicrobials
PS Disclosure: Page 115; 182pp; English.
CC The invention relates to peptides enterocin 900, brochocin-C that inhibit
```

CC bacterial growth. Microorganisms expressing the above peptides are used
 CC to inhibit bacterial growth in foods (specifically meat), live animals
 CC (applied topically), food preparation areas and fermentation vessels. The
 CC bacteriocin enterocin 900, brochoicin-C may be used directly, specifically
 CC to inhibit bacteria that cause mastitis. Antibodies specific for the
 CC peptides are used to detect expression of bacteriocins in cells. Secretory
 CC vectors (containing sequences for a promoter, a leucocin A processing
 CC peptide, a heterologous peptide, lacZ and lacI) are used to secrete
 CC bacteriocins, enzymes or other proteins, e.g. for use in food production
 CC and as probiotics. L. gelidum, or other leucocin A-expressing bacteria,
 CC inhibit the growth of *Listeria* and the sulphide-producing spoilage
 CC organism *Lactobacillus sakei*; they also improve the sensory properties
 CC (smell, taste and colour) of meat, specifically ground beef. L. gelidum
 CC can also be used in preparation of animal feeds (silage), as probiotic,
 CC to control *Salmonella* in poultry intestines and therapeutically against
 CC mucosal tissue infections. Treatment with bacteriocins provides vacuum-
 CC packed meat products with predictable and longer storage life. The
 CC secretion vectors can express several bacteriocins for broader spectrum
 CC of activity, e.g. tailored for a particular target organism.
 SQ Sequence 184 BP; 60 A; 24 C; 37 G; 63 T;

Query Match 52.5%; Score 21; DB 1; Length 184;

Best Local Similarity 73.0%; Pred. No. 29;

Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 TGCTGAATTCGATGCTGATTCGATTCGAAACAA 37
 |||||
 DB 121 TGCTGAATTCGATGCTGATTCGATTCGAAACAA 157

RESULT 11

ID 085373 standard; cDNA: 593 BP.

AC 085373;

DT 29-AUG-1995 (first entry)

DE Growth factor-activated fic gene.

KW Growth factor: fic gene; heparanase; heparin; heparan sulfate;

OS Arthritis; restenosis; cancer; wound healing; ss.

PN Homo sapiens.

PD WO9504158-A.

PN 09-FEB-1995.

PF 26-JUL-1994: U08207.

PR 29-JUL-1993: US-099866.

PR 13-OCT-1993: US-136117.

PA (UPJO) UPJOHN CO.

PI Hoogwerf AJ, Ledbetter SR;

DR WPI: 95-082239/11.

DR P-PSDB: R70803.

PT Screening for cpds. with anti-heparanase activity - by detecting

PT inhibition of heparin or heparan sulphate degradation, cancer.

PT potentially useful for treating arthritis, restenosis, cancer.

PS Disclosure; Page 53; 60pp; English.

CC Purified heparanases, prepared under reducing conditions and

CC activated with transglutaminase, are given in R70786-805. Most

CC are prepared by reverse transcription of mRNA from activated human

CC leukocytes, then cloning of the cDNA into pVL1392 baculovirus

CC vector, and expression in Sf9 cells in the presence of reduced

CC glutathione and dithiothreitol.

SQ Sequence 593 BP; 170 A; 136 C; 135 G; 152 T;

Query Match 52.0%; Score 20.8; DB 1; Length 593;

Best Local Similarity 78.1%; Pred. No. 38;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 9 AATCTAATGCTGATTCATGAAACAAATGA 40

DB 124 ATCCACATGCTGATGTCAGAAACAAAGA 155

RESULT 12

ID T28960/c

ID T28960 standard; cDNA to mRNA: 2859 BP.
 AC T28960;
 DT 28-NOV-1996 (first entry)
 DE Human 26S proteasome constitutive protein p12 cDNA.
 KW Human; 26S proteasome; constitutive protein; p12; treatment;
 OS Homo sapiens.
 FH Key
 FT Location/Qualifiers
 FT cds 1..2859
 FT /tag= a
 FT /note= "STOP codon absent"
 PN J08116972-A.
 PD 14-MAY-1996.
 PF 28-OCT-1994; 264810.
 PR 28-OCT-1994; JP-264810.
 PA (KYOM) KYOMA HAKKO KOGYO KK.
 PI Tanaka K, Yamasaki M, Yano K;
 DR WPI: 96-280771/29.
 DR P-PSDB: R97284.
 PT Human 26S proteasome constitutive proteins p12 and p21 - useful to
 PT diagnose and treat conditions associated with human 26S proteasome
 PT related diseases, e.g. Alzheimer's
 PS Claim 3; Pages 9-10; 12pp; Japanese.
 CC The present sequence encodes the human 26S proteasome constitutive
 CC protein p12, which can be used to treat and diagnose 26S
 CC proteasome associated conditions, e.g. Alzheimer's disease, and in
 CC the study and elucidation of human 26S proteasome function. A cDNA
 CC fragment was synthesised by RT-PCR (science 241, 708-712 (1988)),
 CC on the basis of an amino acid sequence corresponding to the bovine
 CC proteasome constitutive protein p12. The cDNA fragment was used
 CC as a plaque hybridisation probe to isolate phages contg. human
 CC p12 protein DNA, isolated from a human hepatic, or HepG cell line
 CC (ATCC HB8065) cDNA library. The phages were introduced into hosts
 CC by the in vitro packaging method. From the resulting transformants
 CC the protein was isolated and purified.
 SQ Sequence 2859 BP; 875 A; 558 C; 672 G; 754 T;

Query Match 52.0%; Score 20.8; DB 1; Length 2859;

Best Local Similarity 70.0%; Pred. No. 44;

Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 1 TGCTGAATTCGATGCTGATTCGATTCGAAACAAATGA 40
 |||||
 DB 1272 TGCCATTAACTGATGCTGCTTTTTCATGACCCCTTATGA 1233

RESULT 13

T58840.1/c

Continuation (2 of 6) of T58840 from base 100001 (Mycoplasma genitalium genome.)

WP Sequence split into 6 fragments LOCUS T58840 Accession T58840

WP Fragment Name Begin End

WP T58840_0 1 110000

WP T58840_1 100001 210000

WP T58840_2 200001 310000

WP T58840_3 300001 410000

WP T58840_4 400001 510000

WP T58840_5 500001 580073

Query Match 52.0%; Score 20.8; DB 1; Length 110000;

Best Local Similarity 78.1%; Pred. No. 60;

Matches 25; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 7 AATTCGATGCTGATTCATGAAACAAAT 38

DB 33361 AATTCGAACTGATGCTGAAATTAAT 33330

RESULT 14

V28681/c

AC V28681 standard; cDNA: 786 BP.

DT 29-JUL-1998 (first entry)
 DE Ripening banana pulp cDNA clone U-0128 SEQ ID NO:39.
 KW Banana; ripening; pulp; Musa acuminata cv. Grand Nain; fruit;
 KW genetic control; tissue senescence; ss.
 OS Musa acuminata.
 PN WO9811228-A2.
 PD 19-MAR-1998.
 PF 08-SEP-1997; G02424.
 PR 25-APR-1997; GB-008366.
 PR 10-SEP-1996; GB-018862.
 PA (ZENE) ZENECA LTD.
 PI Bird CR, Medina-Suarez Rdf, Seymour GB;
 DR WPI: 98-207389/18.
 PT Modulation of ripening or tissue senescence in bananas - comprises
 PT use of DNA isolated from ripening banana pulp to produce genetically
 PT modified fruit
 PS Claim 1; Page 42-43; 72pp; English.
 CC The present sequence represents a cDNA clone isolated from ripening
 CC banana pulp. 57 clones were isolated and are given in V28643 to V28699.
 CC The cDNA clone sequences can be used in a method of modulating ripening
 CC or tissue senescence process in plants of the genus Musa. The method
 CC comprises: (a) inserting into the plant material at least 1 of the 57
 CC sequences (as above); (b) regenerating the plant material, and (c)
 CC selecting from the transformed regenerants, plants with modulated
 CC ripening or tissue senescence characteristics. Also described in the
 CC present invention are: (1) plants, their progeny, seed and material
 CC obtained from the plants, produced by the above method; (2) a vector
 CC functional in plants comprising a promoter region which is operably in
 CC plant cells, a polynucleotide sequence as defined above, and a
 CC transcription termination sequence; and (3) a method of controlling
 CC plant pathogens comprising the application of anti-pathogenic agent to
 CC plants of (1).
 SQ Sequence 786 BP; 223 A; 186 C; 142 G; 213 T;

Query Match 52.0%; Score 20.8; DB 1; Length 786;
 Best Local Similarity 88.0%; Pred. No. 39;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 7 AATTCTAATGCTGCTATTTCATGA 31
 ||||| ||||| ||||| ||||| ||
 Db 685 AATTCTAATGCTGCTATTTCAGGA 661

RESULT 15
 V70887/c
 ID V70887 standard; cDNA; 4810 BP.
 AC V70887;
 DT 01-MAR-1999 (first entry)
 DE Plasmid pTK-fus-lacZ.
 KW E. coli; vector; gene delivery; gene therapy; fusulin; promoter;
 KW pTK-fus-lacZ; ds.
 OS Chimeric - Melolontha melolontha entomopoxvirus.
 PN WO9850571-A1.
 PD 12-NOV-1998.
 PF 07-MAY-1998; U09326.
 PR 07-MAY-1997; US-852629.
 PA (UYFL) UNIV FLORIDA.
 PI LI YIHALL RL, Moyer R;
 DR WPI: 99-034733/03.
 PT Recombinant entomopoxvirus vector containing an early or late gene
 PT promoter - useful for expressing heterologous proteins in vertebrate
 PT cells
 PS Example 2; Page 31-34; 56pp; English.
 CC This is the nucleotide sequence of plasmid pTK-fus-lacZ, which
 CC contains the lacZ gene under control of the Melolontha melolontha
 CC entomopoxvirus (EPV) early fusulin gene promoter. The plasmid was
 CC used in the construction of recombinant EPV vectors. The invention
 CC provides a recombinant EPV comprising early or late EPV gene
 CC promoters, an early vaccinia virus promoter or a promoter
 CC recognised by vertebrate transcription factors and operatively
 CC linked to a heterologous gene, which allows expression of the
 CC heterologous gene when infected into a vertebrate cell. The EPV

CC does not replicate in the vertebrate cell, but enters the vertebrate
 CC cell cytoplasm and achieves expression of the heterologous gene
 CC without causing cytopathic or cytotoxic effects. It is used as a
 CC gene delivery vector for gene therapy.
 SQ Sequence 4810 BP; 1407 A; 993 C; 1022 G; 1388 T;

Query Match 51.5%; Score 20.6; DB 1; Length 4810;
 Best Local Similarity 74.3%; Pred. No. 54;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 TGTAATCTAATGCTGCTATTTCATGAACAAT 38
 ||||| ||||| ||||| ||||| |||||
 Db 984 TTATTTCTAATGCTGCTATTTCATTCATTAAT 950

Search completed: September 13, 1999, 16:22:01
 Job time: 4021 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:13:55 ; Search time 849.94 Seconds
(without alignments)
92.832 Million cell updates/sec

Title: US-09-325-095-2122.SEQ
Perfect score: 40
Sequence: 1 TGCTGTAATTCATATGCTGATTTTCATGAACAATGA 40

Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 98626752 residues
Database: EST:*

1: em_est1:.*
2: em_est2:.*
3: em_est3:.*
4: em_est4:.*
5: em_est5:.*
6: em_est6:.*
7: em_est7:.*
8: em_est8:.*
9: em_est9:.*
10: em_est10:.*
11: em_est11:.*
12: em_est12:.*
13: em_est13:.*
14: em_est14:.*
15: em_est15:.*
16: em_est16:.*
17: em_est17:.*
18: em_est18:.*
19: em_est19:.*
20: gb_est1:.*
21: gb_est2:.*
22: gb_est3:.*
23: gb_est4:.*
24: gb_est5:.*
25: gb_est6:.*
26: gb_est7:.*
27: gb_est8:.*
28: gb_est9:.*
29: gb_est10:.*
30: gb_est11:.*
31: gb_est12:.*
32: gb_est13:.*
33: gb_est14:.*
34: gb_est15:.*
35: gb_est16:.*
36: gb_est17:.*
37: gb_est18:.*
38: gb_est19:.*
39: gb_est20:.*
40: gb_est21:.*
41: gb_est22:.*
42: gb_est23:.*
43: gb_est24:.*
44: gb_est25:.*
45: gb_est26:.*
46: gb_est27:.*
47: gb_est28:.*
48: gb_est29:.*
49: gb_est30:.*
50: gb_est31:.*
51: gb_est32:.*
52: em_est20:.*
53: em_est21:.*

54: em_est22:.*
55: em_est23:.*
56: em_est24:.*
57: em_est25:.*
58: em_est26:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	23.6	59.0	275	20	Z41154	Z41154 HSC2PF092 n
2	23.6	59.0	331	22	R37868	R37868 YC98F08.s1
3	23.6	59.0	231	22	R42218	R42218 Yf90a11.s1
4	22.4	56.0	361	40	AA936397	AA936397 c049b03.s
5	22.4	56.0	303	50	AV047057	AV047057 AV047057
6	22.2	55.5	298	39	AA897649	AA897649 o172b08.s
7	22.2	55.5	439	40	AA977508	AA977508 om60e05.s
8	22.2	55.5	554	45	AA866471	AA866471 UI-R-E0-b
9	22	55.0	475	25	N44819	N44819 YV39a10.r1
10	22	55.0	343	31	AA298585	AA298585 EST114216
11	22	55.0	633	49	AI632172	AI632172 t685907.x
12	21.8	54.5	246	49	AV024880	AV024880 AV024880
13	21.6	54.0	280	20	T02774	T02774 011M7 gmbp
14	21.6	54.0	316	27	AA037681	AA037681 zK34q11.x
15	21.6	54.0	586	44	AU040107	AU040107 AU040107
16	21.6	54.0	733	45	AI325034	AI325034 m042b04.x
17	21.6	54.0	434	49	AV024572	AV024572 AV024572
18	21.6	54.0	363	49	AV024632	AV024632 AV024632
19	21.6	54.0	222	50	AV026938	AV026938 AV026938
20	21.6	54.0	375	50	AV027720	AV027720 AV027720
21	21.6	54.0	284	50	AV042239	AV042239 AV042239
22	21.4	53.5	444	35	C29010	C29010 C29010 R1ce
23	21.2	53.0	400	22	R63552	R63552 Y109b02.r1
24	21.2	53.0	462	22	R66182	R66182 Y130d06.r1
25	21.2	53.0	377	22	R79702	R79702 Y185d12.r1
26	21.2	53.0	457	25	N71517	N71517 YW33b09.r1
27	21.2	53.0	694	26	W51807	W51807 zC48b03.r1
28	21.2	53.0	435	26	W74515	W74515 zG56b04.r1
29	21.2	53.0	795	27	Z78407	Z78407 HS278407 Hu
30	21.2	53.0	424	28	AA070699	AA070699 zms9e10.r
31	21.2	53.0	433	28	D44928	D44928 HUMSUPY408
32	21.2	53.0	457	29	AA130382	AA130382 z130d06.r
33	21.2	53.0	425	29	AA150158	AA150158 z142a04.r
34	21.2	53.0	501	29	AA174146	AA174146 zp05a05.r
35	21.2	53.0	487	31	AA313615	AA313615 EST185490
36	21.2	53.0	552	37	AA670930	AA670930 vP91f06.r
37	21.2	53.0	435	42	AI148231	AI148231 oZ81b06.x
38	21.2	53.0	818	45	AI344225	AI344225 tc02b09.x
39	21.2	53.0	807	49	AI654097	AI654097 tv61f12.x
40	21	52.5	322	21	T83404	T83404 yd44c03.r1
41	21	52.5	244	28	C13568	C13568 C13568 C1om
42	21	52.5	413	41	AI041697	AI041697 oy33g11.x
43	21	52.5	587	46	AI437916	AI437916 sa41c11.y
44	21	52.5	324	46	AI440791	AI440791 sa48f05.y
45	21	52.5	466	47	AI495116	AI495116 sb03f06.y

ALIGNMENTS

RESULT 1
Z41154
LOCUS Z41154 275 bp mRNA
DEFINITION HSC2PF092 normalized infant brain cDNA Homo sapiens CDNA clone
ACCESSION Z41154
NID Z41154
VERSION Z41154.1 GI:566902

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Insert Size: 1713
High quality sequence stops: 169 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1713 Std Error: 0.00
Seq primer: Promega -21m13
High quality sequence stop: 169.
Location/Qualifiers

FEATURES
SOURCE
1..231
/organism="Homo sapiens"
/db_xref="GDB:401930"
/db_xref="taxon:9606"
/clone="IMAGE:29583"
/clone_lib="Soares Infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: latmid BA; Site: 1: Not
I; Site 2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5']
ACTGAGAGATTCGCGCCCGCAGCAATTTTCTTTTCTTTT 3';
Double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the latmid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
94 a 24 c 29 g 83 t 1 others

Query Match
Best Local Similarity 59.0%; Score 23.6; DB 22; Length 231;
Matches 29; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 CTGTAATCTAATGCTGATTCATGACAAATGA 40
||||| ||||| ||||| ||||| ||||| |||||
Db 14 CTGGAATAATATAATATTATTATTAAGAAATGA 51

RESULT 4
AA936397/c 361 bp mRNA EST 24-AUG-1998
LOCUS od49h03.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1371317 3',
DEFINITION mRNA sequence.
ACCESSION AA936397
NID 93094315
VERSION AA936397.1 GI:3094315
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 361)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Bonaldo, Ph.D.
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044601.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnl.gov/dbtrp/image/image.html

Insert Length: 985 Std Error: 0.00
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 333.
Location/Qualifiers

FEATURES
SOURCE
1..361
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1371317"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site: 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGACGAGGAGCGCCCGCCCAATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
ORIGIN
98 a 68 c 92 g 103 t

Query Match
Best Local Similarity 81.2%; Score 22.4; DB 40; Length 361;
Matches 26; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 9 ATTCTAATGCTGATTCATGACAAATGA 40
||||| ||||| ||||| ||||| ||||| |||||
Db 297 ATTCTAATGCTGATTCATGACAAATGA 266

RESULT 5
AV047057/c 305 bp mRNA EST 18-MAY-1999
LOCUS AV047057 Mus musculus adult C57BL/6J testis Mus musculus cDNA clone
DEFINITION 170065H12, mRNA sequence.
ACCESSION AV047057
NID 94866722
VERSION AV047057.2 GI:4866722
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 305)
Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Ito, M., Aizawa, K.,
Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,
Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,
Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,
Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y.,
Sugahara, Y., Suzuki, H., Suzuki, R., Tateo, M., Tomaru, Y.,
Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,
Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
RIKEN Mouse ESTs
Unpublished (1999)
TITLE
JOURNAL On Jun 22, 1998 this sequence version replaced gi:3247212.

Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098

OY 1 TGCTGTAATTCATGCTGATTCATGAACA 35
 DB 191 TGATGAATTTGTAATGCTGATTCATGAACA 157

RESULT 8

AA866471 554 bp mRNA EST 05-FEB-1999
 LOCUS UI-R-E0-br-g-08-0-UI.s1 UI-R-E0 Rattus norvegicus cDNA clone
 DEFINITION UI-R-E0-br-g-08-0-UI 3' similar to gb|AA123324|AA123324 mm37f10.r1
 Beddington mouse embryonic region Mus musculus cDNA clone 540139
 5' mRNA sequence.

ACCESSION AA866471
 NID 94229998
 VERSION AA866471.1 GI:4229998
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 TITLE Bernaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT On Mar 16, 1998 this sequence version replaced gi:2961932.

FEATURES
 source Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult 18-day embryo library. cDNA Library Preparation: M. Fatima
 Bernaldo, Ph.D. Clone distribution: clones will be available through
 Research Genetics The following repetitive elements were found in
 this cDNA sequence: 103-170, >AT-richflow_complexity
 Seq primer: M13 Forward

FEATURES Location/Qualifiers

1..554
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-E0-br-g-08-0-UI"
 /clone_1lb="UI-R-E0"
 /dev_stage="embryonic"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker. Site.1: NotI. Site.2: EcoRI. This library
 consists of a mixture of individually tagged normalized
 libraries constructed from 8, 12 and 18-day embryo. The
 tag is a string of 3-5 nucleotides present between the
 Not I site and the oligo-dT track which allows
 identification of the library of origin of a clone within
 the mixture."

BASE COUNT 187 a 89 c 83 g 195 t
 ORIGIN

Query Match 55.5%; Score 22.2; DB 45; Length 554;
 Best Local Similarity 77.1%; Pred. No. 2.2e+02;
 Matches 27; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTGATTCATGAACA 35
 DB 211 TGCTTAATGCTGAATGCAACATTCCTGAACA 245

RESULT 9

MA4819/c
 LOCUS MA4819 475 bp mRNA EST 13-FEB-1996
 DEFINITION yy39a10.r1 Soares melanocyte 2NbhM Homo sapiens cDNA clone
 IMAGE:273594 5' similar to SW:PIIA_BOVIN_P32871
 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOFORM ;,
 mRNA sequence.

ACCESSION MA4819
 NID 91185985
 VERSION MA4819.1 GI:1185985
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 TITLE Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P. and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 JOURNAL
 COMMENT On May 8, 1995 this sequence version replaced gi:800198.

FEATURES
 source Contact: Wilson RK
 Washington University School of Medicine
 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LIND ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: T7
 High quality sequence stop: 402.

FEATURES Location/Qualifiers

1..475
 /organism="Homo sapiens"
 /db_xref="GDB:3883236"
 /db_xref="taxon:9606"
 /clone="IMAGE:273594"
 /clone_1lb="Soares melanocyte 2NbhM"
 /sex="Male"
 /tissue_type="melanocyte"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Vector: pT73D (Pharmacia) with a modified
 polylinker. Site.1: Not I. Site.2: Eco RI. 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTTCACATCTGAAGTGGAGCGCCGACATTTTCTTTTCTTTT 3'],
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library constructed by Bento Soares and
 M. Fatima Bernaldo. RNA from normal foreskin melanocytes
 (FS3/4) was kindly provided by Dr. Anthony P. Albino."

BASE COUNT 170 a 90 c 85 g 128 t
 ORIGIN

Query Match 55.0%; Score 22; DB 25; Length 475;
 Best Local Similarity 73.7%; Pred. No. 2.7e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 1 TGCTGTAATTCATGCTGATTCATGAACAAT 38
 DB 372 TGCTGTAATTCATGCTGATTCATGAATTCAT 335

RESULT 10
 AA298585 343 bp mRNA EST 18-APR-1997
 LOCUS AA298585/c
 DEFINITION EST114216 HSC172 cells II Homo sapiens cDNA 5' end similar to
 phosphatidylinositol 3-kinase, alpha catalytic subunit, mRNA
 sequence.

ACCESSION AA298585
 MID g1950938
 VERSION AA298585.1 GI:1950938
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 343)
 Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weissbrock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Val,C., Clifton,R.A., Cline,T.R., Cotton,M.D., Batle-Hughes,J., Fine,L.D., Flytzergaid,L.M., Fitzhugh,W.M., Fitchman,J.L., Geoghagen,N.S., Glodde,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palmanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudak,D.M., Shirley,R., Small,R.V., Spriggs,T.A., Utecherback,T.R., Weisman,J.F., Li,Y., Bednarek,D.P., Cao,L., Cepede,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Melasner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Pannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., and Venter,J.C.
 TITLE Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
 JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
 MEDLINE 96026280
 COMMENT On Sep 12, 1996 this sequence version replaced g1:1392803.
 Other_ESTs: THC168479
 Contact: Kerlavage, AR
 Bioinformatics
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850 USA
 Tel: 3018699056
 Fax: 3018699423
 Email: akerv@vtgtr.org
 For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
 Seq primer: M13 Reverse.
 FEATURES
 source location/Qualifiers
 1..343
 /organism="Homo sapiens"
 /db_xref="ATCC (inhost):179595"
 /db_xref="taxon:9606"
 /clone_lib="HSCI72 cells II"
 /cell_type="fibroblast"
 /dev_stage="fetal"
 /note="Organ: lung; Vector: pbluescript SK-; Site_1: EcoRI; Site_2: XhoI"
 BASE COUNT 132 a 61 c 64 g 84 t 2 others
 ORIGIN
 Query Match 55.0%; Score 22; DB 31; Length 343;
 Best Local Similarity 73.7%; Pred. No. 3e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Oy 1 TCGTGAATTCATGCTGATTCATGCAACAAT 38
 ||||||||||||||||| | | | | |
 ||||||||||||||||||| | | | | |
 Db 287 TCGTGAATTCATGCTGATTCATGCAAT 250
 RESULT 11
 LOCUS A1632172 633 bp mRNA EST 26-APR-1999
 DEFINITION ts85507.x1 NCI-CGAP GC6 Homo sapiens cDNA clone IMAGE:2238108 3' similar to SW:P11A.BOVIN P32871 PHOSPHATIDYLINOSITOL 3-KINASE CATALYTIC SUBUNIT, ALPHA ISOBFORM ;, mRNA sequence.

ACCESSION A1632172
 MID g4683502
 VERSION A1632172.1 GI:4683502
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 633)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On May 7, 1998 this sequence version replaced g1:3121411.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/HLNT at: www-bio.llnl.gov/bdip/image/image.html
 Seq primer: -40UP from G1bco
 High quality sequence stop: 446.
 FEATURES
 source location/Qualifiers
 1..633
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="15g26.1"
 /clone="IMAGE:2238108"
 /clone_lib="NCI-CGAP-GC6"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /note="Vector: pT7R3D-Pac (Pharmacia) with a modified polylinker; Plasmid DNA from the normalized library NCI-CGAP GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.
 BASE COUNT 169 a 109 c 108 g 246 t 1 others
 ORIGIN
 Query Match 55.0%; Score 22; DB 49; Length 633;
 Best Local Similarity 73.7%; Pred. No. 2.5e+02;
 Matches 28; Conservative 0; Mismatches 10; Indels 0; Gaps 0;
 Oy 1 TCGTGAATTCATGCTGATTCATGCAACAAT 38
 ||||||||||||||||| | | | | |
 ||||||||||||||||||| | | | | |
 Db 220 TCGTGAATTCATGCTGATTCATGCAAT 257
 RESULT 12
 LOCUS AV024880 246 bp mRNA EST 02-JUN-1999
 DEFINITION AV024880 Mus musculus adult C57BL/6J lung Mus musculus cDNA clone 1200008B05, mRNA sequence.
 ACCESSION AV024880
 MID 94801872
 VERSION AV024880.1 GI:4801872
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:19:58 ; Search time 329.34 Seconds

(without alignments)
366,951 Million cell updates/sec

Title: US-09-325-095-2425.SEQ

Sequence: 1 AAGATCAGACAAATGCGCTTCTTATGACATCA 38

Scoring table:

IDENTITY_MUC
679419 segs, 1590154660 residues

Searched:

Database:

GenBank: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_om: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_p11: *
8: gb_p12: *
9: gb_p1: *
10: gb_p12: *
11: gb_p13: *
12: gb_p1: *
13: gb_sts: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_vl: *
18: em_fun: *
19: em_hlg: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_om: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_vl: *
34: gb_hlg1: *
35: gb_hlg2: *
36: gb_in1: *
37: gb_in2: *
38: em_ba1: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_p14: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23.6	62.1	1251	1	SYODESA2 D13779 <i>Synechococcus</i>

2	22.8	60.0	204544	35	AC007677	AC007677 Homo sapi
3	22	57.9	144260	11	AC006299	AC006299 Homo sapi
4	22	57.9	238438	34	AC005046	AC005046 Homo sapi
5	21.8	57.4	2131	12	MMHCL	X66285 M.musculus
6	21.6	56.8	17696	7	SPBC15C4	AL0232270 S.pombe c
7	21.6	56.8	116210	10	HS406C18	AL0233290 Homo DNA
8	21.6	56.8	17251	12	AB015623	AB015623 Mus muscu
9	21.6	56.8	149886	12	MMHCL438N12	AF049850 Mus muscu
10	21.4	56.3	212346	34	HS90G24	AF077409 Arabidops
11	21.2	55.8	68251	8	T24M8	AF024499 Caenorhab
12	21.2	55.8	37380	36	CEFA262	Y09612 B.priorididm
13	21	55.3	2460	1	PLPWRAGG	AF053983 Podospora
14	21	55.3	2626	8	AF053983	AF053983 Podospora
15	21	55.3	90731	11	AC004897	AC004897 Homo sapi
16	21	55.3	5437	37	AF110008	AF110008 MOlgula o
17	21	55.3	137611	42	AC006355	AC006355 Homo sapi
18	20.8	54.7	107848	8	ATAC007019	AC007019 Arabidops
19	20.8	54.7	183386	11	AC005229	AC005229 Homo sapi
20	20.8	54.7	176725	35	AC007386	AC007386 Homo sapi
21	20.8	54.7	80389	37	AC005643	AC005643 Drosophila
22	20.6	54.2	119707	7	PINCEPRG	D17510 Pinus thunb
23	20.6	54.2	161516	35	AC007131	AC007131 Homo sapi
24	20.6	54.2	108320	35	AC007333	AC007333 Homo sapi
25	20.6	54.2	207188	35	AC007389	AC007389 Homo sapi
26	20.6	54.2	41396	36	DMC34F3	AL031583 Drosophila
27	20.4	53.7	14039	1	MTU19364	U19364 Methanobact
28	20.4	53.7	81729	7	AB016874	AB016874 Arabidops
29	20.4	53.7	37672	7	SCVCSMGN	X89633 S.cerevisia
30	20.4	53.7	6319	7	SCYOR291W	275199 S.cerevisia
31	20.4	53.7	6918	7	YSCRC1	D90459 Yeast R1C1
32	20.4	53.7	153472	9	AB020868	AB020868 Homo sapi
33	20.4	53.7	118968	10	HS796F18	AL031291 Human DNA
34	20.4	53.7	123397	11	AC004808	AC004808 Homo sapi
35	20.4	53.7	973	11	HS1RPS14	AF058410 Homo sapi
36	20.4	53.7	339	14	G07818	G07818 human STS c
37	20.4	53.7	169	14	HMSMX3123	L77630 Human chrom
38	20.4	53.7	3417	17	ARLVGPN	X52400 Nigeria Las
39	20.4	53.7	1830	17	LSVNG	K03362 Lassa virus
40	20.4	53.7	144952	34	HS417M14	AL024498 Homo sapi
41	20.4	53.7	193126	42	AC006045	AC006045 Homo sapi
42	20.2	53.2	44185	10	HS069729	U69729 Human Xp22
43	20.2	53.2	180664	34	L49408	AC006706 Caenorhab
44	20.2	53.2	83527	37	L49408	L49408 Drosophila
45	20	52.6	3452	4	AF001076	AF001076 Gallus ga

ALIGNMENTS

RESULT 1
SYODESA2
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS

SYODESA2 1251 bp DNA
Synechococcus sp. desa gene encoding delta 12 desaturase, complete cds.

D13779
GI:488510
D13779.1 GI:488510
delta 12 desaturase.
Synechococcus sp.
Synechococcus sp. (strain:PCG7002) DNA, clone lib:lambd DASH II.
Bacteria; Cyanobacteria; Chroococcales; *Synechococcus*.

1 (bases 1 to 1251)
Sakamoto, T.
Direct Submmission
Submitted (27-NOV-1992) to the DDBJ/EMBL/GenBank databases, Toshio Sakamoto, Graduate University of Advanced Studies, Molecular Biomechanics, Myodaiji, Okazaki, Aichi 444, Japan
(Tel:0564-55-7602, Fax:0564-53-7400)

2 (bases 1 to 1251)
Sakamoto, T.
Unpublished (1994)
3 (sites)
Sakamoto, T., Wada, H., Nishida, I., Ohmori, M. and Murata, N.


```
/clone.lib="RPCI-11 human BAC library"
repeat_region 457..499 /rpt_family="AT_rich"
repeat_region 837..861 /rpt_family="AT_rich"
repeat_region complement(862..1010) /rpt_family="AluYo"
repeat_region complement(1011..1060) /rpt_family="(TAA)n"
repeat_region complement(1061..1182) /rpt_family="AluYo"
repeat_region 1183..1210 /rpt_family="AT_rich"
repeat_region 4144..4273 /rpt_family="L1ME"
repeat_region 5137..5273 /rpt_family="FLAM_C"
repeat_region 5300..5493 /rpt_family="L1MD"
repeat_region 5494..5775 /rpt_family="AluX"
repeat_region 5776..5796 /rpt_family="(TAA)n"
repeat_region 5797..6324 /rpt_family="(TAA)n"
repeat_region 6325..6543 /rpt_family="L1MD"
repeat_region 6544..6756 /rpt_family="MER20"
repeat_region 6763..6816 /rpt_family="L1MD"
repeat_region 6925..7073 /rpt_family="AT_rich"
repeat_region 7086..7101 /rpt_family="AluYo/FRAM"
repeat_region 7339..8052 /rpt_family="L1MD2"
repeat_region 8056..8086 /rpt_family="LTR30"
repeat_region complement(8515..8753) /rpt_family="L1MD2"
repeat_region 8878..8943 /rpt_family="L1MA9"
repeat_region /rpt_family="purine-rich"
repeat_region 8952..9239 /rpt_family="AluSc"
repeat_region 9718..9775 /rpt_family="AT_rich"
repeat_region complement(10580..10601) /rpt_family="(TAAA)n"
repeat_region 11225..11247 /rpt_family="AT_rich"
repeat_region 11509..11949 /rpt_family="MT1H"
repeat_region 11988..12011 /rpt_family="(CA)n"
repeat_region 12097..12548 /rpt_family="MER65A"
repeat_region 13119..13157 /rpt_family="AT_rich"
repeat_region 13216..13240 /rpt_family="AT_rich"
repeat_region 13936..14549 /rpt_family="AT_rich"
repeat_region 14550..14620 /rpt_family="L1M4"
repeat_region 14778..14893 /rpt_family="MT2D"
repeat_region 15007..15035 /rpt_family="L1ME1"
repeat_region complement(15856..18793) /rpt_family="AT_rich"
repeat_region complement(19078..19390) /rpt_family="AluY"

repeat_region complement(19413..19720) /rpt_family="MSTB"
repeat_region 19788..19902 /rpt_family="MER91C"
repeat_region 19996..20035 /rpt_family="AT_rich"
repeat_region complement(20379..20919) /rpt_family="MT1D"
repeat_region complement(20988..20998) /rpt_family="MER5B"
repeat_region 20999..21145 /rpt_family="MER5A"
repeat_region complement(21146..21156) /rpt_family="MER5B"
repeat_region complement(21222..21657) /rpt_family="L2"
repeat_region complement(21729..22469) /rpt_family="L2"
repeat_region 22475..22820 /rpt_family="MADE1"
repeat_region 22869..22762 /rpt_family="MIR"
repeat_region 22953..23917 /rpt_family="L1MA7"
repeat_region complement(24118..24224) /rpt_family="MT1J"
repeat_region 24359..24404 /rpt_family="(CA)n"
repeat_region 24481..24516 /rpt_family="AT_rich"
repeat_region 24965..25007 /rpt_family="AT_rich"
repeat_region 26626..26830 /rpt_family="AT_rich"
repeat_region 27718..27944 /rpt_family="LTR33"
repeat_region /rpt_family="MIR"
repeat_region 28087..28150 /rpt_family="(TA)n"
repeat_region complement(28175..28424) /rpt_family="MIR"
repeat_region 29543..29699 /rpt_family="L2"
repeat_region complement(30930..31229) /rpt_family="AluYa5"
repeat_region 31490..31846 /rpt_family="THE1B"
repeat_region 31992..32064 /rpt_family="MIR"
repeat_region 32425..32532 /rpt_family="MT1H"
repeat_region 33624..34049 /rpt_family="MIR"
repeat_region complement(34212..34412) /rpt_family="MER20"
repeat_region complement(35457..35762) /rpt_family="AluSg"
repeat_region complement(36300..36505) /rpt_family="MT1J"
repeat_region 38065..38184 /rpt_family="L1PB1"
repeat_region complement(38455..38577) /rpt_family="L2"
repeat_region complement(39107..41490) /rpt_family="L1M4"
repeat_region 41504..41861 /rpt_family="MT1A1"
repeat_region complement(41874..42343) /rpt_family="L1M4"
repeat_region complement(42348..42710) /rpt_family="L1MD"
repeat_region 43747..44018 /rpt_family="AluX"
repeat_region 44098..44123
```

```

Query Match Similarity      57.9%: Score 22; DB 34; Length 238438;
Best Local Similarity       83.3%: Pred. No. 24;
Matches      25; Conservative    0; Mismatches      5; Indels      0; Gaps      0

Dy      8 AGAACATGCTAGGCTTCTTTAGCCATC 37
        ||||| ||||| ||||| ||||| ||
Db   19036 AGAAACATGCTTAGGCTTCCTTACCCTC 19007

RESULT      5
MMHC1/c
LOCUS
DEFINITION  M.musculus DNA for Hc1 locus.
ACCESSION   X66285
VERSION     X66285.1 GI:51207
KEYWORDS    Hc1 gene.
SOURCE      house mouse.
ORGANISM    Mus musculus.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS     Kahn,T.
TITLE       Submitted (17-MAY-1992) T. Kahn, German Cancer Res Center, Im
            Neuheimer Feld 242, W-6900 Heidelberg, FRG
JOURNAL
REFERENCE   2 (bases 1 to 2131)
AUTHORS     Kahn,T., Friesl,H., Copeland,N.G., Gilbert,D.J., Jenkins,N.A.,
            Glasmann,L., Kramer,J. and zur Hausen,H.
TITLE       Molecular cloning, analysis, and chromosomal localization of a
            mouse genomic sequence related to the human papillomavirus type 18
            ES region
JOURNAL     Mol. Carcinog. 6 (2), 88-99 (1992)
MEDLINE     93000417
COMMENT     Similarity to the human Papillomavirus type 18 ES region at the DNR
            and putative protein level. Similarity of the putative protein
            with the ORF of Q300 (Wagner, et al., J.Virol. 65:3529-3267(1991))
            RNA overexpressed in mouse skin tumours obtained after carcinogen
            (DMBA) or UV treatment. Amplification of the genomic DNA in the Eb
            cell line and one UV induced mouse skin carcinoma.
FEATURES
Source      location/Qualifiers
            1..2131
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /chromosome="8"
            /cell_line="EB"
            /map="Jund, Jund, linked to Plat, Flg"
            1383..1739
            /note="Hc1 ORF. Protein sequence is in conflict with the
            conceptual translation"
            /codon_start=1
            /protein_id="CAA46991.1"
            /db_xref="PID:e49230"
            /db_xref="PID:g1333929"
            /db_xref="GI:1333929"
            /db_xref="MGJ:MGI:96820"
            /db_xref="RNTGY:RNNRNFEELRASSRACTHSIIITHEVDSIFSCVS"
            /translation="RNTGYACICVYAHVSVCYCICVCMCYACVCEVMCKVCC"
            HMNMNCEHVMELCV"
misc_feature 1472..1691
            /note="E5 similarity"
BASE COUNT  591 a      405 c      447 g      668 t
ORIGIN
Query Match      57.4%: Score 21.8; DB 12; Length 2131;
Best Local Similarity 78.8%: Pred. No. 36;
Matches      26; Conservative    0; Mismatches      7; Indels      0; Gaps      0;

```

DB 447 AATAACCAAAAGTCTAGTTCCTTTTC 415

RESULT 6

SPBC15C4

LOCUS SPBC15C4 17696 bp DNA PLN 27-OCT-1998

DEFINITION S.pombe chromosome II cosmid c15C4.

ACCESSION AL023290

NID g3116143

VERSION AL023290.1 GI:3116143

KEYWORDS ABL1 homologue; amino acid permease; ATP dependent rna helicase; geyan1geranyl transferase; zinc finger.

SOURCE fission yeast.

ORGANISM Schizosaccharomyces pombe

Eukaryota; Fungi; Ascomycota; Archiascomycetes; Schizosaccharomycetales; Schizosaccharomycetaceae;

REFERENCE 1 (bases 1 to 17696)

AUTHORS Lyne, M., Rajandream, M.A., Barrell, B.G., Xiang, Z. and Aves, S.

TITLE Direct Submission

JOURNAL Submitted (05-MAY-1998) European Schizosaccharomyces genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: Barrell@sanger.ac.uk and Department of Biological Sciences, University of Exeter, Perry Road, Exeter EX4 4QG, United Kingdom

COMMENT

Notes:

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.

(URL, http://www.sanger.ac.uk/Projects/S_pombe/)

During 1995 to 1996 about 80% of S. pombe chromosome I was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from The European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.

Protein coding regions (CDS) have been predicted with the help of computer analysis using the Genefinder program in Pombase (an ACCEB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites. CDS are numbered using the following system eg SPBC25H2.01c. SP (S. pombe), B (chromosome 2), C25H2 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid c15C4 is overlapped at the 5' end by cosmid 14C8 and at the 3' end by cosmid 21H7.

FEATURES

Source

Location/Qualifiers

1. .17696

/organism="Schizosaccharomyces pombe"

/strain="972h-"

/db_xref="taxon:4896"

/chromosome="II"

/clone="cosmid c15C4"

/map="11R"

1. .1901

/note="nominal overlap with SPBC14C8 S. pombe chromosome 2"

gene

complement(816..1928)

/gene="SPBC15C4.01c"

complement(join(816..1004,1082..1280,1332..1477,1541..1684,1758..1928))

/gene="SPBC15C4.01c"

/note="SPBC15C4.01c, len:282, SIMILARITY:Homo sapiens, Y103_HUMAN, hypothetical protein KIA0103, (297aa), fasta scores: opt: 390, E():2.9e-25, (33.3% identity, in

210 aa)"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAA1892.1"

/db_xref="PID:e1291650"

/db_xref="PID:g3116144"

/db_xref="GI:3116144"

/db_xref="SPTRMBL:060110"

/translation="MNSILKVPDPONPOEIVALLFSQGEAYAKLGKYNDEIMDYOKYV
IAALTGETVIAKKCNMRLENDREHSPRVEGLGMPLEATASEKDAASYNSKLSDEP
THVYIKRKALIRSMGOTKCEIOGLINLYDTPFNDEANALADIYVSVEAFSAIF
CYEMVYLQFEPRFLRDLDTVYLAQSNATNWEFLKHCBSVEICEYFHQMFI
SKCCQQLERSLRKLRLKRVNKISDIPDETSEVQLIQLSLKSDLPKQPKLK
ALRQC"

complement(1005..1019)

/gene="SPBC15C4.01c"

/note="ttacaattaatag, splice branch and acceptor"

complement(1076..1081)

/gene="SPBC15C4.01c"

/note="gtattt, splice donor sequence"

complement(1281..1296)

/gene="SPBC15C4.01c"

/note="ctaattatgtag, splice branch and acceptor"

complement(1326..1331)

/gene="SPBC15C4.01c"

/note="gtaaagt, splice donor sequence"

complement(1478..1490)

/gene="SPBC15C4.01c"

/note="ctaattcatcag, splice branch and acceptor"

complement(1535..1540)

/gene="SPBC15C4.01c"

/note="gtaaag, splice donor sequence"

complement(1685..1701)

/gene="SPBC15C4.01c"

/note="ttaatttttaattag, splice branch and acceptor"

complement(1752..1757)

/gene="SPBC15C4.01c"

/note="gtaaagt, splice donor sequence"

2961..4994

/gene="SPBC15C4.02"

join(2961..3185,3230..3538,3652..4634,4727..4994)

/gene="SPBC15C4.02, len:594, SIMILARITY:Saccharomyces cerevisiae, YLR253W, Y153_YEAST, hypothetical 65.9 kd protein, (569 aa), fasta scores: opt: 1440, E():0, (44.5% identity in 499 aa)"

/codon_start=1

/product="hypothetical protein"

/protein_id="CAA1893.1"

/db_xref="PID:e1291651"

/db_xref="PID:g3116145"

/db_xref="GI:3116145"

/db_xref="SPTRMBL:060111"

/translation="MNSYVLTQVLYVYNSKRRKYMFWLRSFSSNNRKLFLVHLPL
HLSLVISICFDEPGQELVASDHIOQRNHRKRPVAVASTTLMASVALVDPDVKAG
VSSRRAYRVLAGFLCPSDYKRYGSSVASEEERQLASDCHLCARSRSLVFEENG
IYIKIGQHLAMGYVLPKEWNTNVAIKLQDPCPSLNDIDHLEFVDGKGLDEFEFG
DPIALGVASIAQVAKRLKDSQDVAAVQVOPSVSLSDLSLDTWTFWAFKIKTFEPD
FKLMWLADIERISLPDLDTREKNSLTKREPAHLSTSLVPEVWMSKPRITLMYE
VAGARINDLSFMDHRSISROLVSDTCHITREMTFGGGLHCDPHGNNVILRSKPKN
SPSPRNREIVLIDHGLRDIPHEIQVYANRWLNTINFNEKRLFAKKAANVSDENP
KIFATATGTRPKYKSLKSLPIGPNHEEOKRFISALQKGMLQSIILSLTMRPLTL
LMKNTDLVRSIDENLTKTSGPEKLYLLMARCLRCVYDDKMDSLMNSRSFVWSKVFEG
TYVRLSYFSLSLKXTLAEWTFYVHMYL"

3186..3191

/gene="SPBC15C4.02"

/note="gtattt, splice donor sequence"

3219..3229

/gene="SPBC15C4.02"

/note="ctaatttaag, splice branch and acceptor"

3539..3544

/gene="SPBC15C4.02"

/note="gtaaagt, splice donor sequence"

```

misc_feature      3633..3651
                  /gene="SPBC15C4.02"
                  /note="ctaatgataataatttag, splice branch and acceptor"
misc_feature      4635..4640
                  /gene="SPBC15C4.02"
                  /note="gtaatg, splice donor sequence"
misc_feature      4713..4726
                  /gene="SPBC15C4.02"
                  /note="ttaactgacacag, splice branch and acceptor"
gene              5206..6634
                  /gene="SPBC15C4.03"
                  join(5206..5273,5323..6634)
CDS               5206..5273,5323..6634
                  /note="SPBC15C4.03, len:459, SIMILARITY: Saccharomyces
                  cerevisiae, YOR370C, Q06761, mrg6 protein, (394 aa),
                  fasta scores: opt: 287, E():1.5e-30, (31.2e identity in
                  430 aa)"
                  /codon_start=1
                  /product="putative geranylgeranyl-transferase component,
                  GDP dissociation inhibitor"
                  /protein_id="CAA18894.1"
                  /db_xref="PID:e1291652"
                  /db_xref="PID:g3116146"
                  /db_xref="GI:3116146"
                  /db_xref="SPTREMBL:O60112"
                  /translation="MDPNSYDYIVGTNRNLSAISMNORVLIHDENSFGEI
                  DSLTLDLEQINLEKIKVDSQILINDNSHSPKRFVEQLNKDLPKNGSVIQF
                  HOPEIASSSELVLLSEETIKYKILKPARSPRLNDEMIKVPESRADIFNNKLS
                  LASKRIVMRFKPVSNIADEONQNLKWEKSPFYFLFEVLPQSAIESITIGYGC
                  SLKSDIPETDADLTVALKPFHSGMGDYSLAMYGSELECGPCRSSAGVGTPL
                  GGAIDKIDSKITVLRKPSLTAKKIVSYDGEKPHOOIOORILWYKCGCOOLFQDS
                  FFAALDASLIHSPSISENFGNSVQAKITGSSGCGPCPGTCIWTIKTINDSCNPF
                  NIAIKLCSFGSCDDLEITVQDETINQNLHIDYDITLHANSLEFELIQNTFQLR
                  EGFEDDE"
                  5206..6408
                  /gene="SPBC15C4.03"
                  /note="pfam match to entry PF00996 GDI, GDP dissociation
                  inhibitor"
misc_feature      5274..5379
                  /gene="SPBC15C4.03"
                  /note="gtaatg, splice donor sequence"
misc_feature      5302..5322
                  /gene="SPBC15C4.03"
                  /note="ctgaatgtaataatttag, splice branch and acceptor"
gene              5302..5322
                  /gene="SPBC15C4.03"
                  /note="ctgaatgtaataatttag, splice branch and acceptor"
CDS               5302..5322
                  /gene="SPBC15C4.03"
                  /note="ctgaatgtaataatttag, splice branch and acceptor"
                  /complement(7121..8749)
                  /gene="SPBC15C4.04C"
                  /note="SPBC15C4.04C, len:542, SIMILARITY: Neurospora
                  crassa, O59942, amino acid permease, (551 aa), fasta
                  scores: opt: 1647, E():0, (49.8e identity in 484 aa)"
                  /codon_start=1
                  /product="amino acid permease"
                  /protein_id="CAA18895.1"
                  /db_xref="PID:e1291653"
                  /db_xref="PID:g3116147"
                  /db_xref="GI:3116147"
                  /db_xref="SPTREMBL:O60113"
                  /translation="MDSVSNVSYNEQKENDKEGFSNLKSHVSHSETDEVSND
                  DNLQLELGKPVKFKREFSTWATFSFASISGLFAVTVTYSPILISGAPSAVMWCLLI

```

```

Query Match      56.8%; Score 21.6; DB 7; Length 17696;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

```

```

QY      2 AGATCAGACAACTGCTTCTTTCAGCATC 37
        |||||
Db      6019 AGGATCAGAACTATGTCAGAGATTGTGCGTCA TC 6054

```

```

RESULT 7
HS406C18/c LOCUS      HS406C18      116210 bp      DNA      PRI      12-MAR-1999

```

```

DEFINITION      Human DNA sequence from clone 406C18 on chromosome Xq27.1-27.3.
                  Contains two MAGE (Melanoma Associated Antigen) family pseudogenes,
                  GSSs and an STS, complete sequence.
ACCESSION      AL023773
NID            93449129
VERSION        AL023773.1 GI:3449129
KEYWORDS       HTG; MAGE; Melanoma Associated Antigen.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 116210)
REFERENCE      Pearce, A.
AUTHORS        Direct Submission
TITLE          Submitted (02-SEP-1998) E-mail enquiries: humquery@sanger.ac.uk
JOURNAL        Clone requests: clonerequest@sanger.ac.uk
COMMENT        On Aug 21, 1998 this sequence version replaced gi:3288039.
                  During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
                  together with a note of the overlapping clone name. Note that the
                  variation annotation may not be found in the sequence submission
                  corresponding to the overlapping clone, as we submit sequences with
                  only a small overlap as described above.
                  This sequence is the entire insert of clone 406C18. The true left
                  end of clone 142F18 (AL031073) is at 94113 in this sequence.
                  This sequence has been finished according to sequence map criteria
                  as follows. An attempt is made to resolve all sequencing problems,
                  such as compressions and repeats, but not necessarily within known
                  annotated human repeat sequence elements (e.g. Alu). Where the
                  sequence is ambiguous, there is an annotation using the 'unsure'
                  feature key.
                  This sequence was generated from part of bacterial clone contigs of
                  human chromosome X, constructed by the Sanger Centre Chromosome X
                  Mapping Group. Further information can be found at
                  http://www.sanger.ac.uk/HGP/ChrX
                  406C18 is from the library RPI3 constructed at the Roswell Park
                  Cancer Institute by the group of Pieter de Jong. For further
                  details see http://pacpac.med.buffalo.edu/VECTOR:pcrPAC2.

```

FEATURES

source

```

1..116210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="406C18"
/map="q27.1-27.3"
/clone_11b="RPI3"
/complement(1..157)
/note="THE1B repeat: matches 158. .1 of consensus"
1505..1803
/note="ALUSP repeat: matches 1. .297 of consensus"
3326..3414
/note="MIR repeat: matches 3. .95 of consensus"
/complement(4189..4254)
/note="L1PA8 repeat: matches 910. .846 of consensus"
5239..5306
/note="34 copies 2 mer ta 87% conserved"
/complement(5363..6216)
/note="THE1B-INTERNAL repeat: matches 859. .1 of consensus"
/complement(6217..6568)
/note="THE1B repeat: matches 364. .1 of consensus"
/complement(7285..7660)
/note="THE1B repeat: matches 364. .3 of consensus"
/complement(7663..8822)
/note="THE1B-INTERNAL repeat: matches 1580. .427 of
consensus"

```

```

repeat_region   complement(8839..9195)
                  /note="THE1B-INTERNAL repeat: matches 350. .1 of consensus"
repeat_region   complement(9198..9585)
                  /note="THE1C repeat: matches 369. .1 of consensus"
repeat_region   complement(10000..10678)
                  /note="L1R8 repeat: matches 691. .1 of consensus"
12345..12498
/note="MER4A repeat: matches 1. .146 of consensus"

```

```

repeat_region 12509..12586
/note="MER4A repeat: matches 228..308 of consensus"
12587..12890
/note="MER4A repeat: matches 367..661 of consensus"
repeat_region complement(12882..13148)
/note="MLR1D repeat: matches 287..1 of consensus"
repeat_region complement(13162..13209)
/note="MIR repeat: matches 138..90 of consensus"
13437..15600
/note="L1PA4 repeat: matches 730..893 of consensus"
complement(16025..16313)
/note="AluX repeat: matches 294..3 of consensus"
repeat_region 19307..19348
/note="21 copies 2 mer aa 81% conserved"
19388..19775
/note="MLR1C repeat: matches 58..456 of consensus"
20264..20789
/note="L1MA4 repeat: matches 514..1045 of consensus"
complement(20866..21101)
/note="L1MA1 repeat: matches 1051..614 of consensus"
21096..21220
/note="L1MA2 repeat: matches 667..796 of consensus"
complement(21209..21362)
/note="MER45 repeat: matches 168..11 of consensus"
complement(23167..23470)
/note="MLR1D repeat: matches 455..154 of consensus"
23522..23571
/note="25 copies 2 mer tt 82% conserved"
24349..24442
/note="MIR repeat: matches 80..178 of consensus"
complement(24458..24683)
/note="MIR repeat: matches 256..21 of consensus"
complement(25195..25446)
/note="match: GGG B34150"
25885..25914
/note="15 copies 2 mer tt 87% conserved"
26931..27144
/note="L1PA4 repeat: matches 678..891 of consensus"
27916..28087
/note="MER20 repeat: matches 1..178 of consensus"
28842..28889
/note="24 copies 2 mer gt 100% conserved"
29010..29490
/note="MLR1D repeat: matches 1..505 of consensus"
29816..29924
/note="MIR repeat: matches 1..114 of consensus"
complement(31395..31696)
/note="AluX repeat: matches 302..1 of consensus"
32451..33106
/note="L1PA9 repeat: matches 238..905 of consensus"
33416..33594
/note="MIR repeat: matches 70..252 of consensus"
33740..33927
/note="L1MA7 repeat: matches 846..1034 of consensus"
33997..34095
/note="MER5A repeat: matches 4..109 of consensus"
complement(34339..34686)
/note="MLR1A1 repeat: matches 365..1 of consensus"
complement(35068..35168)
/note="MER5A repeat: matches 189..83 of consensus"
complement(35382..35418)
/note="MIR repeat: matches 151..114 of consensus"
35515..35672
/note="MIR repeat: matches 54..206 of consensus"
36412..36759
/note="L1PA5 repeat: matches 545..892 of consensus"
36983..37010
/note="14 copies 2 mer aa 89% conserved"
complement(37417..37465)
/note="MIR2 repeat: matches 146..95 of consensus"
complement(37765..37929)
/note="MER45 repeat: matches 178..15 of consensus"
complement(38232..38332)

```

```

repeat_region complement(38603..38724)
/note="MIR2 repeat: matches 22..146 of consensus"
repeat_region complement(38763..39144)
/note="L1MC2 repeat: matches 1075..665 of consensus"
repeat_region complement(39256..39911)
/note="L1MA6 repeat: matches 1047..312 of consensus"
complement(40016..40315)
/note="L1MC1 repeat: matches 669..370 of consensus"
40587..41141
/note="L1PA1 repeat: matches 338..896 of consensus"
44248..44562
/note="AluX repeat: matches 1..302 of consensus"
complement(44565..44715)
/note="AluO repeat: matches 281..122 of consensus"
complement(45854..45954)
/note="L1MB3 repeat: matches 338..238 of consensus"
45959..45980
/note="11 copies 2 mer at 100% conserved"
complement(46036..46247)
/note="L1MB5 repeat: matches 218..7 of consensus"
46271..47164
/note="MER21A repeat: matches 25..902 of consensus"
complement(47170..49364)
/note="L1 repeat: matches 3872..1681 of consensus"
49365..50875
/note="L1 repeat: matches 3874..5390 of consensus"
50733..51615
/note="L1PA2 repeat: matches 1..890 of consensus"
51852..51885
complement(52467..53595)
/note="L1 repeat: matches 5223..4076 of consensus"
53604..54173
/note="L1MA10 repeat: matches 334..912 of consensus"
54174..54207
/note="17 copies 2 mer ca 100% conserved"
54240..54388
/note="L1MA9 repeat: matches 902..1049 of consensus"
complement(54850..55704)
/note="MLR2 internal repeat: matches 5214..4354 of consensus"
complement(56831..56959)
/note="MLR2 internal repeat: matches 2337..2406 of consensus"
57082..57127
/note="23 copies 2 mer gt 100% conserved"
complement(57128..57258)
/note="L1ME2 repeat: matches 896..760 of consensus"
complement(57268..57589)
/note="AluX repeat: matches 302..1 of consensus"
complement(57616..58103)
/note="L1MD2 repeat: matches 724..227 of consensus"
complement(58189..58402)

```

```

Query Match 56.8%; Score 21.6; DB 10; Length 116210;
Best Local Similarity 75.0%; Pred. NO. 36;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 1 AAGGATCGAACAATGGCTAGCGTTCTTACCAT 36
DB 60604 AATGCCAGACAGCACTTGGCTTTTACCAT 60569

RESULT 8
LOCUS AB015623 17251 bp DNA ROD 30-JUN-1998
DEFINITION Mus musculus Cyp21 and Slp genes for steroid 21-hydroxylase and
AB015623 sex-limited protein, complete and partial cds.
ACCESSION AB015623
NID 63273213
VERSION AB015623.1 GI:3273213
KEYWORDS X; Tnx; tenascin-X; Slp; sex-limited protein; steroid
21-hydroxylase.

```

SOURCE Mus musculus (strain:Fl animals of C57BL6 and CBA cross)
 cell_line:IT2 DNA.

ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 17251)
 AUTHORS Matsumoto, K. and Ikuta, T.
 TITLE Direct Submission
 JOURNAL Submitted (15-JUN-1998) to the DDBJ/EMBL/GenBank databases.
 Ken-ichi Matsumoto, Graduate School of Pharmaceutical Sciences,
 Hokkaido University, Department of Molecular Biology; Kita 12,
 Nishi 6, Kita-ku, Sapporo, Hokkaido 060-0812, Japan
 (E-mail: kmatsumapharm.hokudai.ac.jp, Tel: 81-11-706-3731,
 Fax: 81-11-706-4988)
 2 (bases 1 to 17251)
 AUTHORS Matsumoto, K. and Ikuta, T.
 TITLE Mus musculus 5' truncated pseudogene of tenascin-X, steroid
 21-hydroxylase (Cyp21), and sex-limited protein (slp) genes,
 partial cds and complete sequences
 JOURNAL Published Only in Database (1998) In press
 COMMENT Interspersed repeats and simple sequence repeats were identified
 with RepeatMasker accessed through
 http://ftp.genome.washington.edu/cgi-bin/RepeatMasker.html.
 FEATURES
 source
 1. 17251
 /organism="Mus musculus"
 /strain="Fl animals of C57BL6 and CBA cross"
 /db_xref="taxon:10090"
 /cell_line="IT2"
 /chromosome="17"
 /note="1. 3793: Genomic PCR product; our used primer set
 is as follows: forward primer corresponds to nt 58711 to
 58735, and reverse primer to nt 63202 to 63222. Each nt
 number is based on the sequence from Accession
 No. AB010266.; 1632..17251: A mouse genomic clone:
 this clone was obtained by the screening with a mouse TNX
 cDNA probe, 800-bp Hind III fragment which encodes FNIII
 repeat M26-M28."
 260..538
 /note="high similarity to nt 58991-59269 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 1229..1511
 /note="high similarity to nt 60069-60350 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 2322..2630
 /note="high similarity to nt 61162-61470 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 2935..3265
 /note="high similarity to nt 61780-62115 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 3516..3638
 /note="high similarity to nt 62365-62487 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 3737..3877
 /note="high similarity to nt 63145-63288 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 3967..4097
 /note="high similarity to nt 63378-63508 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 4220..4352
 /note="high similarity to nt 63631-63763 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 4446..4597
 /note="high similarity to nt 63862-64013 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 4698..4794
 /note="high similarity to nt 64114-64210 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 4882..5043
 /note="high similarity to nt 64297-64458 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 5125..5288
 /note="high similarity to nt 64540-64703 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 complement(5441..6120)
 /gene="Cyp21"
 /number=10
 complement(5441..8490)
 /gene="Cyp21"
 complement(5441..5446)
 /gene="Cyp21"
 /note="This polyA signal is used for Cyp21 gene."
 5590..5845
 /gene="Cyp21"
 /note="high similarity to nt 65005-65293 of mouse
 tenascin-X (GenBank Accession Number AB010266)"
 complement(join(5852..6120,6205..6308,6378..6547,
 6744..6944,7058..7144,7233..7334,7421..7516,7622..7764,
 8125..8214,8279..8480))
 /gene="Cyp21"
 /note="similar to mouse steroid 21-hydroxylase A (GenBank
 Accession Number M15009) or mouse steroid 21-hydroxylase B
 (GenBank Accession Number M15008)"
 /codon_start=1
 /product="steroid 21-hydroxylase"
 /protein_id="BAA31153.1"
 /db_xref="PID:d1032100"
 /db_xref="PID:g3373214"
 /db_xref="GI:3273214"
 /translation="MLPGLLLLLLAGTRWLGQWKLRLKLLPPLAPGFLHFLQPN
 LPILYLGTLKGLPIYRIILGLMDVVVLSNRTIEALLQKWDVAFAPHLNKLMDL
 DLSIGDYSLMKAKRLSRSLMGLMDSMEPLLEQLQECERRAQAQVPIHNEF
 EFLPLGSLISGLTFGDKDSTLYORLHDCVODLLQANWMSIQILITLPLRFLNPF
 OKLQIOESRPHIYKQDLRKDKSLVAGKMDVYLQCYEKORQDGRDLHEGV
 HMSVDFPIGTEITATLLSNVAFLLHHPETQRLDEDLKIGPSQLIRRMKL
 PLNATIAEVLRLRVPVLAIPHATSSISGYDIKDWVILIPNIGANLDEWVEL
 PSKMPREFLEGNRPSPSCARVCLGEPRLARLFLVYLAFLALQAFLLPDPGT
 LPSLPQPYAGINLPIPFVYRLQPLAPDQGERP"
 complement(6205..6308)
 /gene="Cyp21"
 /number=9
 complement(6378..6547)
 /gene="Cyp21"
 /number=8
 complement(6744..6944)
 /gene="Cyp21"
 /number=7
 complement(7058..7144)
 /gene="Cyp21"
 /number=6
 complement(7233..7334)
 /gene="Cyp21"
 /number=5
 complement(7421..7516)
 /gene="Cyp21"
 /number=4
 complement(7622..7764)
 /gene="Cyp21"
 /number=3
 7829..7922
 /rpt_family="B1-F"
 7916..8010
 /note="purine-rich"
 8068..8097
 /note="POLX_G"
 complement(8125..8214)
 /gene="Cyp21"
 /number=2
 complement(8279..8490)
 /gene="Cyp21"
 /number=1
 8438..8479
 /rpt_unit="CAG
 complement(8963..9096)
 /rpt_family="PB1"
 complement(9170..9346)
 exon
 gene
 polyA_signal
 misc_feature
 CDS
 exon
 exon
 exon
 exon
 exon
 exon
 exon
 exon
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 exon
 exon
 repeat_region
 repeat_region
 repeat_region
 repeat_region
 exon
 repeat_region
 repeat_region
 repeat_region
 repeat_region

```

repeat_region /rpt_family="B4"
9348..9538 /rpt_family="B3"
repeat_region complement(9580..9922)
/Note="HEVR"
repeat_region complement(10266..10296)
/Note="POLY_A"
repeat_region complement(10409..10445)
/rpt_unit="CAA"
repeat_region complement(10499..10835)
/Note="MMVL30"
repeat_region complement(10839..11036)
/rpt_unit="GAA"
repeat_region complement(11062..11191)
/rpt_family="PB1D7"
11226..11295 /rpt_family="PB1D7"
repeat_region /rpt_family="B4A"
11242..11346 /rpt_family="PB1D9"
11352..11415 /rpt_family="ID5"
repeat_region complement(11583..11649)
/rpt_family="MIR"
11685..11735 /Note="A-rich"
repeat_region 12189..12281
/rpt_family="PB1D10"
12315..12327 /rpt_family="B4"
repeat_region 12328..12385
/rpt_family="ID3"
12386..12423 /rpt_family="B4"
exon complement(13200..13440)
/Note="high similarity to nt 18060-18300 of mouse Slp gene
(GenBank Accession Number M64933)"
polyA_signal complement(13216..13221)
/Note="This polyA signal is used for SLP gene."
gene complement(13299..16983)
/Note="SLP"
CDS complement(join(13299..13440,13577..13709,13994..14077,
14212..14310,14388..14477,14599..14701,15215..15289,
15419..15509,15604..15784,15871..15964,16924..16983))
/Note="SLP"
/Note="codon_start=1"
/Note="product='sex-limited protein'"
Query Match 56.8%; Score 21.6; DB 12; Length 17251;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
OY 1 AAGGATCAGACATGCTGAGCTTCTTTCATGCCAT 36
DB 976 AAGTACAGAAATTTCCATAGCTTCTGATGCCAT 941

```

RESULT 9
LOCUS NMHC438N12/c
DEFINITION Mus musculus major histocompatibility locus class III region:
complement C4 (C4) and cytochrome P450 hydroxylase A (CYP210H-A)
genes, complete cds; slp pseudogene, complete sequence; NG6, SKI,
and complement factor B (Bf) genes, complete cds; and complement
factor C2 (C2) gene, partial cds.
ACCESSION AF049850
NID AF049850
VERSION AF049850.1 GI:2944418
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1. (bases 1 to 149886)

```

AUTHORS Rowen, L., Qin, S., Lasky, S. R., Loretz, C., Dors, M., Mahairas, G. and
Hood, L.
TITLE Sequence of the mouse major histocompatibility locus class III
region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 149886)
AUTHORS Rowen, L.
TITLE Direct Submission
JOURNAL Submitted (23-FEB-1998) Department of Molecular Biotechnology, Box
357730 University of Washington, Seattle, Washington 98195, USA
COMMENT This sequence overlaps BAC 29N7, found in GenBank Accession Number
AF030001, by 93 bases on the 3' side of 29N7. Sequencing
methodology: high redundancy shotgun. Interspersed Repeats were
identified with RepeatMasker (available from
http://ftp.genome.washington.edu/RM/RepeatMasker.html) Simple
sequence repeats were identified with spunik (available from
http://www.geospiza.com/people/chris/software/spunik.html).
FEATURES
source Location/Qualifiers
1..149886
/organism="Mus musculus"
/strain="129"
/db_xref="taxon:10090"
/chromosome="17"
/clone_lib="Genome Systems BAC library"
/haplotype="H2-b"
1..93
/Note="Overlap with BAC 29N7, GenBank Accession Number
AF030001, at positions 201873-201964"
1..20297
/Note="this span duplicates the region between 80946 and
99794; it contains the C4 gene and a copy of the 3' end of
the G7 gene, exons 5-7"
1..3119
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="BAC 232L6"
/sub_clone="PCR product"
complement(27..313)
/rpt_family="MMVL30"
complement(316..423)
/rpt_family="(GGA)n"
370..450
/Note="complex (CT)n microsatellite repeat -- low quality
data"
complement(414..543)
/rpt_family="PB1D7"
596..700
/rpt_family="PB1D9"
706..763
/rpt_family="ID4"
939..971
/rpt_family="POLY_A"
1941..1998
/rpt_family="PB1D10"
2899..149886
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="BAC 438N12"
complement(2992..18332)
/Note="C4"
/Note="complement C4 gene"
complement(join(2992..3133,3271..3403,3688..3771,
3906..4004,4082..4171,4310..4412,4926..5000,5130..5220,
5315..5495,5582..5675,6676..6735,7230..7397,7479..7711,
7906..8077,8166..8282,8378..8534,8638..8713,8872..9081,
9207..9296,9383..9434,9910..10113,10242..10381,
10571..10641,10746..10860,10953..11147,11401..11475,
12139..12465,13349..13507,14161..14346,14442..14627,
14773..14949,15035..15150,15468..15600,15676..15781,
16297..16393,16494..16576,16670..16958,17038..17108,
17317..17518,17975..18167,18268..18332))
/Note="C4"
/Note="intron-exon boundaries defined in relation to cDNAs

```

```
in M1729, M1789"
/codon_start-1
/product="complement C4"
/protein_id="AAC05279.1"
/db_xref="PID:92944420"
/db_xref="GI:2944420"
/translation="NRLLGLAVFSPFCASLQKPKLLILFSPSVNLGTPISGVOLL
DAPGQEVKGSVFLNPKGSGSPKDFKLSDDVLSLEPLEDVNSGCFELDKR
APHIQVADSPMLRNTAFKATETQGNVLFSRRGHIETQDPIINPGQRNRYFA
LDQMRPSTDELTIVENSGLVLRKEITSTISIDAFTIPDISEPGKMSAFS
DGLSNRSTHFEVKKYLPNFEVRIIPMKRYILMVSNSDEIOLDIOARIYKPGVG
VATRFALMDEQGRTEFLNGLETQALVEGRTHISIKQFOFALDKINIGVDELG
RLYATAVIESPCGMEPAELMSRVSASFSDLSRTYRHVPGAHKINIGVQMS
GSEASNPVKVSKTIVSGDSOYLDIOQSNIGQVSIPIPIPTTELELVSASL
YPAIRLVQAPSRGTLIEPLDRSPSVDTILNCPGIPAPFAPFHYVMII
SRQINAMGRPERKTVSVLVLDHOLAPSFYVAFFYHGHVANSLLINOSRCE
GKIQLVYDAKEXERNADMMKLRIOTDSKALVALGAVDMLYAVGSRHPLDMSKYE
VINSYVNGGPGGGDALQYFDAGLAFSDGRLTQTREDLSCPREKSKOKKNVFO
KAVSEKIGOYSSPDARCCQDGMTKIPMKRTGQBARAPQOCREPIFSCCFADL
RNRQTSQAHNLNHNMLQEBDLIEDDLVTSFPEMLMVEEVDLSKLTVLP
DSMTWEINGVSLSKSGLCVAKPTVVRFRKHLRLPLISIRREQELRVLVNYI
LNDQVAVSVHTPEVGLCLAGGMAQOQVTPAGSARPAFVSVPAAANVPKVAR
GYEDLDVASKIQLIEKEGRIHREELVYNDLPNNIGRTLEIGSSDPNIVPGDFS
LVATASEPLETSEGALSPGVASLRLPOGCAOTMTIAPLTASNYLDREOM
SKLSPETKHANDLIQGYMRIOQFRKNDGSRGAMHRDSSMTLFPVLIKILSIAEO
VGNSPERKLOETASMLAOGLGDSFDPVPIRANQGLVGSDEVALTAFVYVILH
HGDLVFODDAKOLKNRVEASITKANSFLOKNSAGLGLHAAAITAVALTAKASD
LNRVANSLMAAEETGEHLVWGLVGSODKVVLRPTAPSPTEPPOAPALMIEETA
YALHLILREGKGMADKASWLTGQSEFGARSTQDVITLDAISAVIASHTEE
KALNVLSSGMRNGLKHGHLNHNHVGLEELKESLGSSTISVYEGSKGTLKLR
TYNVLMKNTTODDQILEVYKVAVERAMANDYEDYDMPAADPSVLPVPILO
LPEGRSRRRRERAPRYVERESVQTVTCIMRNGKGLSGMAIADITLISGRFALAD
LEKJLSDRVYSHFETDGHVLLFYDSPVTRCEVGFEGSDSEVGLVQPSAVLYD
TYSYPRDKCSYFVAAPKSQLATLCSDVOCOCQCPRLRLRERVRDCKGRFRF
TCYPRVEYGFYVLRQEDRAFLRFESKITQVLFHRDTMASICQTNFTLSRASC
LRLEPNKXELIMDGETSDNKDPOYLLDSNTWIEWMSEQCKSTRRHACFOQLKD
FLMEFSSRGCOV"
complement(6866..7037)
repeat_region
  /rpt_family="URR1A"
  9638..9759
repeat_region
  /rpt_family="PB1D7"
  9678..9759
repeat_region
  /rpt_family="PB1D9"
  11554..11584
repeat_region
  /rpt_family="(CA)n"
  11585..11981
repeat_region
  /rpt_family="MTR"
  11979..12096
repeat_region
  /rpt_family="(CA)n"
  12463..12505
repeat_region
  /rpt_family="(GGA)n"
  12506..12621
repeat_region
  /rpt_family="(GGCA)n"
  12609..12727
repeat_region
  /rpt_family="(GAA)n"
  12735..12782
repeat_region
  /rpt_family="(GCA)n"
  12797..12916
repeat_region
  /rpt_family="(GGA)n"
  12845..12962
repeat_region
  /rpt_family="(GGA)n"
  12982..13104
repeat_region
  /rpt_family="(GGA)n"
  13141..13259
repeat_region
  /rpt_family="(GGA)n"
  13262..13318
repeat_region
  /rpt_family="(GGA)n"
  13551..13820
repeat_region
  /rpt_family="B4A"
  14019..14137
repeat_region
  /rpt_family="(CA)n"
  15814..16084
repeat_region
  /rpt_family="B4"
```

```
repeat_region 16086..16200
  /rpt_family="B1-F"
repeat_region 17661..17775
  /rpt_family="PB1D7"
repeat_region 17816..17902
  /rpt_family="PB1D7"
repeat_region complement(18595..18693)
  /rpt_family="(CA)n"
repeat_region complement(20122..20306)
  /rpt_family="B3"
misc_feature 20304..22376
  /note="this span duplicates the region from 105857-107756;
  it contains a copy of exons 3 and 4 of the G11 gene"
repeat_region complement(20314..20457)
  /rpt_family="B1-MM"
repeat_region complement(20461..20663)
  /rpt_family="B4A"
repeat_region complement(21345..21431)
  /rpt_family="(GGA)n"
repeat_region 21913..22022
  /rpt_family="PB1D9"
repeat_region 22126..22222
  /rpt_family="PB1D10"
repeat_region complement(22379..22721)
  /rpt_family="ORR1A3"
repeat_region 22788..22924
  /rpt_family="B3"
repeat_region 22938..23864
  /rpt_family="RLTR13C"
repeat_region complement(23896..23934)
  /rpt_family="(GAAA)n"
repeat_region 25028..26295
  /rpt_family="NME1N"
repeat_region 26296..26443
  /rpt_family="B1-MM"
repeat_region 26445..26523
  /rpt_family="POLY_A"
repeat_region complement(26524..26782)
  /rpt_family="RLTR1"
```

Query Match 56.8%; Score 21.6; DB 12; Length 149886;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 27; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGATCAGAACATGCTAGGCTTCTTAGCCAT 36
DB 71409 AAGTACCAGATTTCCATAGTCTCTGAGCAT 71374

RESULT 10
HS90G24 212346 bp DNA HTG 11-JUN-1999
LOCUS Homo sapiens chromosome 22 clone 90G24, WORKING DRAFT SEQUENCE, in
DEFINITION unordered pieces.
ACCESSION AL008723
NID 95002638
VERSION AL008723.7 GI:5002638
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 212346)
AUTHORS Mashreghi-Mohammadi M.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerryesanger.ac.uk Clone requests: clonequest@esanger.ac.uk
On Jun 7, 1999 this sequence version replaced gi:4902702.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be

contaminated with foreign sequence from E.coli, yeast, vector, phage etc. Order of segments is not known; 800 n's separate segments. Unfinished: dj90G24 Contig_ID: 05038 acc-AL008723 Length: 2325 bp Unfinished: dj90G24 Contig_ID: 04825 acc-AL008723 Length: 4169 bp Unfinished: dj90G24 Contig_ID: 04564 acc-AL008723 Length: 1015 bp Unfinished: dj90G24 Contig_ID: 04865 acc-AL008723 Length: 4623 bp Unfinished: dj90G24 Contig_ID: 04788 acc-AL008723 Length: 3973 bp Unfinished: dj90G24 Contig_ID: 05022 acc-AL008723 Length: 1609 bp Unfinished: dj90G24 Contig_ID: 00436 acc-AL008723 Length: 2078 bp Unfinished: dj90G24 Contig_ID: 04713 acc-AL008723 Length: 1701 bp Unfinished: dj90G24 Contig_ID: 04934 acc-AL008723 Length: 9778 bp Unfinished: dj90G24 Contig_ID: 04936 acc-AL008723 Length: 1537 bp Unfinished: dj90G24 Contig_ID: 05387 acc-AL008723 Length: 2260 bp Unfinished: dj90G24 Contig_ID: 04325 acc-AL008723 Length: 168478 bp.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will be preserved.

FEATURES

BASE COUNT 55209 a 45984 c 45491 g 56862 t 8800 others

ORIGIN

Query Match 56.38; Score 21.4; DB 34; Length 212346;
Best Local Similarity 80.68; Pred. No. 42;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

7 CAGACATGCGCTTCTTACCATC 37
||||||| ||||| ||||| ||||| |||||
Db 105398 CAGACATGCGCTTCTTACCATC 105428

RESULT 11
LOCUS T24M8 68251 bp DNA PLN 04-DEC-1998
DEFINITION Arabidopsis thaliana BAC T24M8.
ACCESSION AF077409
NID G3319365
VERSION AF077409.1 GI:3319365
KEYWORDS
SOURCE
ORGANISM

thale cress.
Arabidopsis thaliana
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

1 (bases 1 to 68251)
Washington University Genome Sequencing Center.
The A. thaliana Genome Sequencing Project

REFERENCE
JOURNAL
AUTHORS
TITLE
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Submitted (10-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
4 (bases 1 to 68251)
Waterston, R.
Direct Submission
Submitted (03-AUG-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:

Genome Sequencing Center

Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, Washu, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is T24H24, 200 bp overlap; 3' cosmid is T25H8.
Actual start of this cosmid is at base position 72886 of T24H24;
actual end is at 68251 of T24M8.

NOTES:

Coding sequences below are predicted from computer analysis, using
the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

SOURCE

1. 68251
Location/Qualifiers

/organism="Arabidopsis thaliana"

/cultivar="Columbia"

/db_xref="taxon:3702"

/clone="T24M8"

/chromosome="IV"

/map="unknown"

/complement(467..2435)

/gene="T24M8.2"

/complement(join(467..713,832..1734,1843..2435))

/gene="T24M8.2"

/note="similar to maize transposon MDR mudra protein
(GB:AL021710)"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAC28217.1"

/db_xref="PDB:G3319368"

/db_xref="GI:3319368"

/translation="MFRVSLFQMPRLTSLIIOACVETSSIFDQSSAFSTGLSD
ADSLFGRKRDQENFTLMFRVSKHSEFFHVKSDLTIVLHCDENCKMLRAK
AGSESEVIRKIVSHSCDSSLRNVSRRQAFPLRLISNHLGKGLPLGROLLET
FRDHDGIVYSKAMRVOEHVELARGLPDSEFVLRPGYKRLKRIISIDGALHTSF
KGTILGASADQGNENLPSAFATVDEENDSMPFLKCLINIIPEDELDVSDRAS
IASGLSENYPLAHGLCTFHLQKLEHFGSSLIIPVYASAVYRTEDSFMT
NSDKLAQVLYEVDVRKMSRAYSPSNRYENSTNLASVALLKQNEYVIVCLEST
RSIMTRFENRERREBSOHPSAVITIVGKKKASIDSTRLEVCQVQEEFEVKGDK
TFLVNDKRTCTCMEDIDKFPYAHGJASKKHILNKNMVFDEFHSTYRNQAVSEI
HPNGDMYEWEIPDTVSEVICLPSTRTVPSRRRRKRIPVWEHGRSQPKRLHKSCRC
GSGHNSKTCVAA1"

6349..8207

/note="pseudogene"

/complement(16052..20499)

/gene="T24M8.3"

/complement(join(16052..16549,16634..16718,16801..16994,
17108..17161,17278..17361,17508..17588,17713..17831,
17882..17954,18017..18097,18226..18339,19111..19260,
19333..19446,19551..19794,19838..20049,20149..20499))

/gene="T24M8.3"

/note="contains similarity to human CD44R1 gene product"

```

(CB:X56794)"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28218.1"
/db_xref="PID:93319369"
/db_xref="GI:3319369"
/translation="MADIDKFLSPMGRESFLKAVGRPKRNIGKSTGRQLTID
PKLSVLSQLOQNTFLKGFPAQLQIAFNINIGLDVLDVDSSTILMKLTHPKQ
VLSLPQIALLESNLKLPVDPILVGDLMASQSWAEDDEKRVAMVDLIVSHQFQ
KIDMVGDAITLPPISVYKRGKRGHGPARYSTRSPKELNLSDEKSRILSELQKVE
EJSNRWMLKLEKAKTKYRFRKSTKSSSFVACSSISGRKKIMEVPIOSQTSSEHLP
VYQPEVAVQVDPDDHLDKTODSDSPLINTEITDDPMQVFPVTPLOSSESNQDNG
NVPYDPYKQDQANEDVDSDQSDSERPTTHSGDLPKRENSSELETNAETVDC
KQODSLDRTASHSHIDLPIDQSSSEQAKINSELIMVFKQVNLMSIEFNAANNIT
LCLSHIIFYVNIQIDTMAAASHLGFAPAKKRVLSAVALSQLPTEPVND
DQTKMAYVYSHPIDTNSKPKNKOLESPPTEDQITDSODIGADLDMQDEGVVD
SDSSPAREREKPTLSEAEVLYAELLSEKTSYELLPSSKSEFALFENITASAPNT
HLTSSGYLISNKPILLSAKPTNMVSTLMEVLSLSQKLATLTLLNORAAYVQWFA
NHLOKSKSFKAAMKSRKNSPKQFVYGNTEPADIDITTYIPMTDSKRWYGLA
NLGVMSVEILDPTDLIDEDEVRKFIIEVYVIMPLIQRICKPCCSONHGLQPYTK
RLDGLYKRLRSDCGPVAM"
complement(23787..26319)
/gene="T24M8.4"
complement(join(23787..23958,25017..25198,25452..25499,
25812..25917,26126..26319))
/gene="T24M8.4"
/notes="contains similarity to C3HC4-type zinc fingers
(pfam: zf-C3HC4.hmm, score: 32.94)"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28219.1"
/db_xref="PID:93319370"
/db_xref="GI:3319370"
/translation="MTLESYADVHTVPOSENSMEERGGLFVPLNFAVNDGIFR
SGFPPVSFSFQSLKLSIYLCEAVPEVREFAKSGIQVFGELRCVRLVEP
FVNIDPEVYREALOHRGTGLVGCVRKIQWCLSTIDEYORFAAKARISDRPELEF
DISNLKHTPLSFSCSRGVNNDSCVCCMRKKAFLPCGHTFCVCSREYWLNRGS
CPLCNRPILIEIDY"
complement(32881..34474)
/gene="T24M8.5"
complement(join(32881..33058,33121..33174,33274..33499,
33597..33661,33782..33824,34028..34171,34424..34474))
/gene="T24M8.5"
/notes="similar to chloroplast triose phosphate
translocators"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28223.1"
/db_xref="PID:93319374"
/db_xref="GI:3319374"
/translation="MISSIRPVLPBLTAIVGIGIYFAIMWALNGVFNYYNKKVLAAP
YLMLTLTSLAGCSLMLVSVVALAHTIGHVAIRCALAAMELNFNMNGMGMIS
NLAFFRNIFSKGKMKGSVYNTYACLSMNSLITVFPFASVAGPOMWADQNDV
SKSDOTLSKMWAVSHVYHLVNOVSYPRLCNHMLPMLKRVNLAIAIILGFIY
SOIKRVRKNNHLLVLCIGMLEPIVITL"
complement(38759..40368)
/gene="T24M8.7"
complement(join(38759..38977,39025..39081,39131..39242,
39720..39951,39999..40140,40198..40368))
/gene="T24M8.7"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28222.1"
/db_xref="PID:93319373"
/db_xref="GI:3319373"
/translation="MNFYKDKKVIISGKHEGLYLEGVYVAAKESVPAENGSGSV
KAVGVRRKSKRESVQGDSESTFESDCLSEYSEKSEGPSHNPILLRDNKMMF
CDKRGIVYEMKELEKASRIQGMENRNRKAKVYLLALDPSQEKATKMTATVGT
NITSQIENLNRREGIGRNLPRRSWCEYCKRRRGQKTLVREGSLYSLVVLG
WINTQNLKSLSLFESTLKLGLSVAFAGSHSSNSLILSLKGLHLHVENRKR
WGKETMADGEEFVCSPEVYGVWCMCFHDMKR"
complement(41605..43570)
/gene="T24M8.6"

```

CDS

```

complement(join(41605..42323,42398..42665,42947..43570))
/gene="T24M8.6"
/notes="Similar to pectinesterase; T24M8.6"
/codon_start-1
/evidence-not_experimental
/protein_id="AAC28220.1"
/db_xref="PID:93319371"
/db_xref="GI:3319371"
/translation="MLVVESEFFILMTMVIGVSKYECODKHSCONFLELKTASSS
LSEIRRDLILIVKNSYRKIDMAMIGMEDTKOHEEMENDLCKEDTNLFEDMES
AKDRIIRVEILLGGEPPYIGSYNHNHTWLSGLTSYTICDELGDGAKRVPRLO
DLISRAKALALFISISRDNTELVNSVPNSPLSHYDKDLILNBOALKIADIV
VAKQGTGKTNVNAIAPRPHSKRFTITIKTIGIVAEIVAEVTKPMTLILIGQDS
TITIGNSASVNRRTFYATPDNSNFKNMEASNGGFGVDMCFRTVGPAPGAV
ALRVSGDMSVYRREVEGYQDALPHIDRQFRCFTGTVDYFCGMAAVFCQIY
ARONMGQSNFTIKOSRETKDKSGFSIONCNITASSDLPATYKTLGRWRIFSTY
AVLOSFEILDVDPGMMWEGEETLSLTHREYQNRGAVTSIRVAMSGFKVMKDK
QATPEYVAKLIDGETWLKESRIPKSGI"
complement(47434..56353)
/gene="T24M8.8"
complement(join(47434..47887,53313..56353))
/gene="T24M8.8"

```

gene

CDS

```

Query Match          55.8%; Score 21.2; DB 8; Length 68251;
Best Local Similarity 76.5%; Pred. No. 54;
Matches 26; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
QY      5 ATCAGACAAATGCTTAGCGTTCCTTACCATCA 38
DB 18646 AACACAGAAATCAAGCCTTCTTAACATCA 18613
      1 ||||| | ||||| |||||

```

RESULT 12

```

CELLF42G2/C      CELF42G2  37380 bp  DNA      INV      16-SEP-1997
LOCUS             Caenorhabditis elegans cosmid F42G2.
DEFINITION        AFO244499
ACCESSION          AF024449
VERSION            AF024449.1 GI:23944467
KEYWORDS
SOURCE
ORGANISM           Caenorhabditis elegans strain-Bristol N2.
                  Caenorhabditis elegans

```

```

REFERENCE
AUTHORS            Bonfield,J., Burton,J., Connell,M., Cogsey,T., Cooper,J.,
                  Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
                  Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
                  Johnston,L., Jones,M., Kersey,J., Kirsten,J., Laister,N.,
                  Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B.,
                  O'Callaghan,M., Parsons,J., Percy,C., Rickert,L., Roopra,A.,
                  Saunders,D., Showkhen,R., Smalton,N., Smith,A., Sonhammer,E.,
                  Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaundin,M.,
                  Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
                  Wilkinson-Sproat,J. and Wohlman,P.

```

TITLE

Nature 368 (6466), 32-38 (1994)

JOURNAL

MEDLINE

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

COMMENT

Submitted by: Direct Submission
Submitted (11-SEP-1997) Department of Genetics, Washington University, St. Louis, Missouri 63108, USA
Submitted by: Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and

Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: twenematode.wustl.edu and jes@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is F30D4, 200 bp overlap. Actual start of this cosmid is at base position 1 of CELF42G2; actual end is at 37407 of CELF42G2.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

SOURCE

1. 37380

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="II"

/clone="F42G2"

/complement(508..6271)

/gene="F42G2.4" /complement(join(508..777,833..1484,1530..1648,1721..1884,2233..2446,2499..2768,3118..3543,5021..5200,5246..5864,6162..6271))

/gene="F42G2.4"

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70356.1"

/db_xref="pid:g2394470"

/db_xref="gi:2394470"

/translation="MNVSELSFSEMOQIVAPTLINLPLOYAEVLEKLVGTDLASAKVC

RGLSAYDKLPLPNSIRFQWMDVRRHNGIEIKTPKPVDTIYAYNGLEKATP

GENETRAFKDLGELKHAITELSLYNRDILNFEFEALRTNKPVIKSLRYDES

INGLEKLYQFDVLEDELEFEVSNNGFEKIRSLDMKNASKFTENGCAFDIKEIE

HLFHEKILITANKFTVDAVKIRDDILKSTFECSTYFEGYNSNPRIKIKVOPDR

TEGDNLMHSGSKERLHFAVHMFSCIKINSYAVNLCOATNEMILITKLDVRSCHGV

SOIRDKGELTIFEDKSAFDVKEYEDGADTVYREVKLVHGXEDPDVACEDNHI

LTKATLAEDLSFLIYATNDNVAADRVHQRKSLQKQKGPVATYTLKLDNIS

KLRVLOQFDPTLTLELINDTNRHKEITIEGDIYEDMDKNKKEINLNFVSIP

FEHFSCHESAVVEWKSISADIIKLIEDFNAPERRIIVHYKSLDYKRLKTLFEG

TYDMNRVSIINNREBQCTLANOYIAHPYGFSDONSRSRNAPFLSYKLAHID

SCRCTROKFSATNLVAGRTTLMETLTKVKLCONEAIAAGRLSLSTSGPTM

LHICVINGKTQYDVKHLYLRVRSALIRQCIDIKDPVAVATTIYGGTSRVEI

FGKDSKLFYISGSKSDQTKLDISNFSNMLLIKQKATLIELEIYDEERKES

EHLTIECEEFVPLDRLSEVEKFTVNRNPEIIVGQLTYMDPVSLRIEIGSEVNR

PWKQIEIDNKLINLNMENHAKKELTINNVKCDVEYLHIFSEVDIILDMKLONV

YLKDTFLSPFCFKQFVNSDINETIIFVDTFLGDPSSRVSDSSGLPNIWFESYAD

VDEDLQIYYOERQFMETRIIPRETRFQHFHRL"

/complement(7207..8767)

/gene="F42G2.5" /complement(join(7207..7250,7300..7600,7776..7830,

7885..8003,8052..8197,8245..8535,8599..8767))

/gene="F42G2.5"

/note="similar to Aplysia californica vesicle-associated

membrane protein/synaptobrevin binding protein

(GB:U06779); also contains similarity to C3HC4-type zinc

misc-feature

gene

CDS

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70357.1"

/db_xref="pid:g2394471"

/db_xref="gi:2394471"

/translation="MPEKSLLEVPPEVSGPDPHYITTYMTLNTISQSPVCLVKT

TPVNSYCRPNRKGILNTGDSKQIEVMQLPEKVPDPDKQKRLVQSCVAPSDVDE

SVKRVKPEDELYNRVLTVDNDNSKSNICEICTLEFSSVNOIPIRIQCHSIC

OSCAKSLAKNCVILPCPCRNFTVSVASLPKRNALLOAIPOKLRVPLEIARQO

LSCOLSKLEOVSKTDSIKAVGAOLSNLKGREPOKRSIVSYASTRETINALE

NEAGIKLHTIDANVSKNALLVASTELIKQNKIKIEIKSMEDVDELATRSLLTS

GTECAVVPANVLDVLTPEIDRMALSMITTFPSNSTV"

/complement(10285..10355,10407..11300,11351..11465)

/gene="F42G2.3"

/complement(7207..8767)

/gene="F42G2.3"

/codon_start=1

/evidence-not_experimental

/protein_id="AAB70355.1"

/db_xref="pid:g2394469"

/db_xref="gi:2394469"

/translation="MNSKPLSYPSIGSLIOHLKPKKFLSQOCCRTTTSIDKSPLY

LKSLKLNESYKLGTYTELNTIENSILKPRAPFDPVDSQSIDLMEVPOEOEEL

VDDRNDPTSTNEORISRADIKNAPSSYTHYIRFVGNHSEQLYERKRLDMKYL

IDKLGGSPPTIKARQFCVENSGLIRPRHLKIRAKVQNLIRKRLNATQHLIDIK

SLPLSGVILSDPLNANPDLIDFLOTAKEMLLCYEGLLVRHRRVINYALYD

ATFLFEDMHEGREGVTHFSFARNENGIIDVADKIRINYNVEGTAEQNLTHINN

FTDIVSHKVDVTLKRYQGRV"

/complement(117255..20720)

/gene="F42G2.6"

/complement(join(17255..17415,17644..18898,18958..19040,

19094..19147,19772..20720))

/gene="F42G2.6"

/note="contains similarity to homeobox domains; coded for

by C. elegans cdna yk74c12.5; coded for by C. elegans cdna

yk51c3.5; coded for by C. elegans cdna yk20c12.5; coded

for by C. elegans cdna yk20c12.3; coded for by C. elegans

cdna yk51c3.3; coded for by C. elegans cdna yk74c12.3

/codon_start=1

/protein_id="AAB70358.1"

/db_xref="pid:g2394472"

/db_xref="gi:2394472"

/translation="MPTSSPTPIYEDLEPTIYQDCYCHRELYFCNKKMGGEITFK

NSGTSNISRVOQFCCVYKQDEIDEMSEKVFYKGLDPRVQNNCEQLSPLS

KFCBHYGMIEFQFNNSLKDVTYADVIDKQDGCCEFHADVLTINLIFEKSH

RDEYGOIICDIDIMMHTREHGKMSIFNSQGFEEYEMNRRVAAQCFKLMQDVTTEY

EVADRTOGCEPESGIESLNPONLKVPLDSKFLSALIVPSPASLVCGRFSPALFPO

KILAEIVRKEIFRPPDPKIRIMLFILOKTLTIRLRIDKSNSSPSYATIHLEDO

EYRTKTISSDCKRKISIKTHLEDELFKRRARRARRSKRANSENEDITSQYQTKL

TKMGTEGYFLRNKHEYFKITLVTVGSSGVKRCVTEKLTSEYQIIDQETWRKFEEN

SRVAIDAEAREQVHRQKLNLSNKRREEAENDEKLEKEMAELEIONVRIIPNSSD

OISRAATLAEEYGTASNIKSRVNSVLSAISTQORLAKYNKVPNGLVYCGEIL
 TSECKEKVNIDFEPKPIINTSLYCDNKFTEALDELSDQKFGFTVMDGNALFG
 TLGNTFDIDVHKFVSVDLPKRGQGSALRPAREKRNRYRVAELAVONFITND
 KVNVMGIYLAGSADFNDAIDMDQSLATKVIKVDVSYGNGENQALIELASEIL
 GNVKFIQEKILIGYFEELISQDTGICVGEDTLKALELAGAVETLIVPENLEVRWL
 KDSNGAELIISTROODANRMDKETGOEMEVVQSESLFWIAEHYKDFGTTLEF
 VSDRSTEGNCKFKFGGIGILRKVNFQDLNEVDNDDEYTD*

BASE COUNT 597 a 802 c 616 g 611 t
 ORIGIN

Query Match 55.3%: Score 21; DB 8; Length 2626;
 Best Local Similarity 73.0%: Pred. No. 76;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AGGATCAGACAAATGCTAGGCTTCTTACCATCA 38
 ||| ||||| || ||| ||||| ||||| ||
 Db 977 AGGTCAGAAAAGACATACGTATCTTCAGCCACCA 941

RESULT 15
 AC004897 90731 bp DNA PRI 14-JAN-1999
 LOCUS AC004897 Homo sapiens clone D0811N16, complete sequence.
 DEFINITION AC004897
 ACCESSION AC004897
 NID 94156184
 VERSION AC004897.1 GI:4156184
 KEYWORDS HTG.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 AUTHORS Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 TITLE 1 (bases 1 to 90731)
 JOURNAL Waterston, R.H.
 REFERENCE The sequence of Homo sapiens clone
 AUTHORS Unpublished
 TITLE 2 (bases 1 to 90731)
 JOURNAL Waterston, R.H.
 REFERENCE Direct Submission
 AUTHORS Submitted (12-JUN-1998) Genome Sequencing Center, Washington
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 JOURNAL MO 63108, USA

REFERENCE 3 (bases 1 to 90731)
 AUTHORS Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (14-JAN-1999) Genome Sequencing Center, Washington
 TITLE University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 JOURNAL MO 63108, USA

COMMENT On Jan 14, 1999 this sequence version replaced gi:3213100.
 FEATURES Location/Qualifiers
 source 1. 90731
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="D0811N16"

BASE COUNT 24311 a 21133 c 20819 g 24468 t
 ORIGIN

Query Match 55.3%: Score 21; DB 11; Length 90731;
 Best Local Similarity 73.0%: Pred. No. 64;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

vY 1 AAGATCAGAACAAATGCTAGGCTTCTTACCATC 37
 ||| ||| ||||| ||| ||||| ||||| ||
 Db 72100 AAGGACAGAGCAATGCTAAGATTCTTTAAATC 72136

Search completed: September 13, 1999, 16:20:17
 Job time: 4529 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:22:01 ; Search time 133.08 Seconds

(without alignments)
71.440 Million cell updates/sec

Title: us-09-325-095-2425_SEQ

Perfect score: 38
Sequence: 1 AAGGATCAGAACATGCGCTTTCATTAGCCATCA 38

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	52.6	658	1	Hepatocyte growth
2	19.4	51.1	4000	1	Mannose-1-phosphat
3	19.4	51.1	1078	1	N. excelsior water
4	19.2	50.5	1863	1	Human FRX1 gene. R
5	19	50.0	3412	1	Human p110 CDNA. R
6	19	50.0	3207	1	p110 CDNA. Recombi
7	19	50.0	3498	1	Platins 3-kinase 11
8	18.8	49.5	1332	1	Enterococcus faeca
9	18.6	48.9	1731	1	Human papilloma vi
10	18.6	48.9	900	1	staphylococcus aur
11	18.6	48.9	58407	1	Methanococcus jann
12	18.6	48.9	3276	1	Human secreted pro
13	18.6	48.9	2933	1	Human secreted pro
14	18.6	48.9	6234	1	Staphylococcus aur
15	18.4	48.4	12923	1	Sequence of human
16	18.4	48.4	5186	1	E. coli PPC gene
17	18.2	47.9	134525	1	Total base sequenc
18	18.2	47.9	9100	1	H. influenzae dete
19	18.2	47.9	2088	1	Coltineau fruit spec
20	18.2	47.9	110000	1	Continuation (9 of
21	18.2	47.4	1581	1	Catalase gene. Hum
22	18	47.4	320	1	Human genome fragm
23	18	47.4	3159	1	Arg-gingipain-2 ge
24	18	47.4	7266	1	Porphyrinomas ging
25	18	47.4	3094	1	Arg-gingipain-1 ge
26	18	47.4	3071	1	Human chromosome-1
27	18	47.4	8640	1	P. gingivalis haem
28	18	47.4	1296	1	H. pylori cell env
29	18	47.4	7265	1	Arg-gingipain high
30	18	47.4	3159	1	DNA encoding Arg s
31	18	47.4	2282	1	Nucleotide sequenc
32	18	47.4	8640	1	Haemagglutinin pro
33	18	47.4	373	1	EST clone C2598. N
34	18	47.4	4403	1	Staphylococcus aur
35	18	47.4	6482	1	Staphylococcus aur
36	18	47.4	8519	1	Enterococcus faeca
37	18	47.4	6422	1	Polynucleotide seq
38	18	47.4	45546	1	Human kidney amino
39	18	47.4	2264	1	Catalase gene of h
40	17.8	46.8	540	1	Human Flap gene. N
41	17.8	46.8	540	1	Sequence encoding
42	17.8	46.8	14985	1	Human box-dependen
43	17.8	46.8	5035	1	Full-length PANCA

ALIGNMENTS

C 44	17.8	46.8	14985	1	V15693	Human Blnl gene, B
C 45	17.8	46.8	387	1	V36214	cDNA sequence of t
ALIGNMENTS						
RESULT	1					
T78411						
AC	T78411	standard; cDNA; 658 BP.				
DE	15-OCT-1997	(first entry)				
KW	Hepatocyte growth factor variant 2.2kb HGF/NK1 terminal portion cDNA.					
OS	HGF; mitogenesis; proliferative disorder; Kringe domain; inhibitor;					
PH	cancer; transplant; toxin; chemotherapeutic; human; ss.					
FT	Key	Location/Qualifiers				
FT	cds	1..39				
FT		/*tag= a				
FT	primer_bind	44..62				
FT		/*tag= b				
FT	primer_bind	621..640				
FT		/*tag= c				
FT	misc_feature	522..528				
FT		/*tag= d				
FT		/*note= "Undetermined sequence"				
PN	MO9640914-A1.					
PD	19-DEC-1996					
PR	07-JUN-1996; U08698.					
PR	07-JUN-1995; US-484841.					
PA	(AARO/) AARONSON S A.					
PA	(BOT/) BOTARO D P.					
PA	(CHAN/) CHAN A M L.					
PA	(CIOC/) CIOCE V.					
PA	(RUB/) RUBIN J S.					
PA	(STAH/) STAHL S J.					
PI	(WING/) WINGFIELD P T.					
PI	Aaronson SJ, Bottaro DP,					
PI	Stahl SJ, Wingfield PT;					
DR	WPI: 97-099934/09.					
DR	P-PSDB: W23714.					
PT	Truncated hepatocyte growth factor (HGF) variants - used to inhibit					
PT	HGF-induced mitogenesis, useful in treating proliferative disorders					
PS	Example 2; Fig 10; 89p; English.					
CC	The present sequence encodes the terminal portion of the coding sequence					
CC	and part of the 3'-untranslated region of the truncated hepatocyte					
CC	growth factor (HGF) 2.2kb HGF/NK1. HGF/NK1 is a substantially pure					
CC	truncated HGF variant comprising the N-terminal and first Kringe domain					
CC	of HGF. HGF/NK1 may be used in cell culture methods to stimulate ex vivo					
CC	mitogenesis in cells expressing the HGF, such cells can be transplanted					
CC	into a subject in need of such cells. HGF/NK2, which is a truncated form					
CC	of HGF comprising the N-terminal and the first two Kringe domains, is					
CC	used to inhibit HGF-induced mitogenesis in cells expressing the HGF					
CC	receptor (HGF4), and it is useful in treating hyper-proliferative					
CC	disorders including cancer and non-malignant disorders. Both HGF/NK1					
CC	and HGF/NK2 are non-toxic and bind specifically to HGFr and may					
CC	therefore be used as carriers for other compounds, e.g. permitting					
CC	the evaluation of the usefulness of certain toxins as chemotherapeutic					
CC	agents.					
SQ	Sequence 658 BP; 216 A; 126 C; 126 G; 180 T;					
Query Match						
Best local similarity 52.6%; Score 20; DB 1; Length 658;						
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;						
QY	1 AAGGATCAGAACATGCGCTTTCATTAGCCAT 36					
DB	358 AAGGATGTAAGAGCTTAGCCTTTCATTAGCAT 393					
RESULT	2					

```

ID T91902
AC T91902 standard: DNA: 4000 BP.
DT 07-MAY-1998 (first entry)
DE Mannose-1-phosphate transferase gene MNNA.
KW Mannose-1-phosphate transfer; MNNA gene; enzyme; yeast; regulatory gene;
KW human; high mannose type neutral saccharide chain; ss.
OS Saccharomyces cerevisiae.
FH Key Location/Qualifiers
FT CDS 418..3953
FT FT /*tag= a
FT FT /product= MNNA
PN J09266792-A.
PD 14-OCT-1997.
PE 29-MAR-1996; 075667.
PR 29-MAR-1996; JP-075667.
PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
DR WPI; 97-553460/51.
DR P-PSDB; M30763.
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast -
PT useful for high mannose type neutral saccharide chain production
PS Claim 2: Page 14-17, 23pp; Japanese.
CC This sequence represents the gene of the invention, designated MNNA,
CC encoding a protein which positively regulates mannose-1-phosphate
CC transfer in yeast. The gene is useful for the preparation of human high
CC mannose type neutral saccharide chain.
SO Sequence 4000 BP; 1453 A; 750 C; 770 G; 1027 T;

Query Match 51.1%; Score 19.4; DB 1; Length 4000;
Best Local Similarity 70.3%; Pred. No. 20;
Matches 26; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

UY 1 AAGGATCAGAACATGCGCTAGCGCTTCTTTATCCATC 37
   1 | | | | | | | | | | | | | | | | | | | |
DB 1861 AATGATCAGAACATTCACAAAGCTTACTTAAGCAAC 1897

RESULT 3
V56043/C
AC V56043 standard: cDNA: 1078 BP.
DT 09-DEC-1998 (first entry)
DE N. excelsior water channel protein, Nem1p1 encoding cDNA.
KW Nicotiana; cell-membrane localised water channel protein; plant;
KW water potential; plant activity; Nem1p1; Nem1p2; Nem1p3;
KW water stress resistance; ds.
KW Nicotiana excelsior.
FH Key Location/Qualifiers
FT CDS 34..897
FT FT /*tag= a
FT FT /product= "cell-membrane localised water channel
FT protein Nem1p1"
PN J10229883-A.
PD 02-SEP-1998.
PE 24-FEB-1997; 039163.
PR 24-FEB-1997; JP-039163.
PA (NISR ) JAPAN TOBACCO INC.
DR WPI; 98-524287/45.
DR P-PSDB; W80568.
PT Improving water potential and maintaining plant activity - comprises
PT introducing plant cell membrane-localised water channel protein gene
PT into plant, produces plants resistant to water stress
PS Examples; Page 6; 11pp; Japanese.
CC This cDNA encodes a Nicotiana excelsior cell-membrane localised water
CC channel protein, Nem1p1. This can be used in a method for improving
CC water potential and maintaining plant activity. The method comprises
CC introducing a plant cell membrane-localised water channel protein gene,
CC especially the N. excelsior Nem1p1, Nem1p2 and Nem1p3 genes, into a
CC plant. The process can be used to produce plants with resistance to
CC water stress.
SO Sequence 1078 BP; 261 A; 228 C; 247 G; 342 T;

```

```

Query Match Similarity      51.1%; Score 19.4; DB 1, Length 1078;
Best Local Similarity      70.3%; Pred. No. 15;
Matches      26; Conservative      0; Mismatches      11; Indels      0; Gaps      0

OY      2 AGGATCAGAACAAATGCTAGGCTTCTTAGGCATCA 38
      ||||| | | | | | | | | | | | | | | | |
DB      660 AGGACATGAGAACTCTAGCATTTCTCTTGACATCA 624

RESULT      4
T47705/c
ID      T47705 standard; cDNA; 1863 BP.
AC      T47705;
DT      20-MAY-1997 (first entry)
DE      Human FRX1 gene.
KW      FRX1; FRX2; FRX1; Fragile X syndrome; mental retardation;
OS      diagnosis; transgenic mouse; knockout mouse; ds.
FH      Homo sapiens.
FT      Key
FT      primer_bind
FT      Location/Qualifiers
FT      complement (1777..1798)
FT      /*tag=
FT      /note="primer used in FRX1 gene chromosome
FT      mapping"
FT      PN
FT      PD      WO9638467-A1.
FT      PD      05-DEC-1996.
FT      PR      31-MAY-1996; U08853.
FT      PR      31-MAY-1995; US-455073.
FT      PA      (TYPE-) UNIV PENNSYLVANIA.
FT      PI      Dreyfuss G, Sloni MC, Zhang Y;
FT      P-PSDB; W09429.
PT      New fragile X related genes FRX1 and FRX2 - used to develop prods.
PT      for identifying individuals with mental retardation caused by
PT      fragile X syndrome
PS      Claim 2; Page 57-59, 75pp; English.
CC      cDNA clones (T47705 and T47706) respectively code for novel human
CC      proteins FRX1 (W09429) and FRX2 (W09430), which are related to FMRI,
CC      the protein associated with Fragile X syndrome. To obtain the human
CC      FRX1 cDNA, a Xenopus laevis ovary cDNA library was screened with
CC      human FMRI cDNA to obtain FMRI and FRX1 (for FMRI unreacting
CC      relative) cDNAs. The Xenopus FRX1 cDNA was then used to isolate
CC      human FRX1 cDNA from a HeLa cell library. The FRX1 gene is
CC      autosomal (unlike FMRI) and is located at 12q13. The FRX1 and FRX2
CC      genes can be used to produce FRX1 and FRX2 recombinant polypeptides,
CC      to develop FRX1 and FRX2 knockout mice, and to develop methods of
CC      screening individuals for FMRI deficiency.
SO      Sequence      1863 BP;      620 A;      320 C;      470 G;      453 T;

Query Match
Best Local Similarity      50.5%; Score 19.2; DB 1, Length 1863;
Matches      24; Conservative      0; Mismatches      8; Indels      0; Gaps      0

OY      1 AAGGATCAGAACAAATGCTAGGCTTCTTAG 32
      ||||| | | | | | | | | | | | | | | | |
DB      1050 AAGACCTGACACATTTCCAAATGCTTCTTAG 1019

RESULT      5
Q51156/c
ID      Q51156 standard; cDNA; 3412 BP.
AC      Q51156;
DT      12-APR-1994 (first entry)
DE      Human p110 cDNA.
KW      phosphoinositide kinase; PI; p85 subunit; screening; agonist;
KW      antagonist; cell proliferation; inhibition; prophylaxis; therapy;
KW      platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
KW      ds.
OS      Human.
FT      Key
FT      Location/Qualifiers
FT      key
FT      cds
FT      /*tag= a
FT      1..3207
FT      17

```


FT /note= "PI3- kinase p110"

PN M09321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parkerj, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R43342.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PT activity, useful for controlling cell proliferation

PS Claim 7; Fig 16; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (anti)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also 051155 and 057522-3.

SO Sequence 3412 BP; 1128 A; 616 C; 706 G; 962 T;

Query Match 50.0%; Score 19; DB 1; Length 3412;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 AGCGTTCTTTAGCCATCA 38

DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 6

ID 051155 standard; cDNA; 3207 BP.

AC 051155.

DT 12-APR-1994 (first entry)

DE p110 cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KM ss.

FT Key Location/Qualifiers

FT cds 1..3207

FT /*tag= a

FT /*note= "PI3- kinase p110"

PN M09321328-A.

PD 28-OCT-1993.

PR 13-APR-1993; G00761.

PR 13-APR-1992; GB-008135.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI Parkerj, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

DR P-PSDB: R43341.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PT activity, useful for controlling cell proliferation

PS Claim 7; Fig 9; 146pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity, and is useful for

CC screening for (anti)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC prophylaxis or therapy. Platelet or neutrophil activity or blood

CC glucose levels can be controlled using the kinase.

CC See also 051155, 059012-23 and 057522-3.

SO Sequence 3207 BP; 1028 A; 581 C; 680 G; 918 T;

Query Match 50.0%; Score 19; DB 1; Length 3207;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 AGCGTTCTTTAGCCATCA 38

DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 7

ID 057012 standard; cDNA to mRNA; 3498 BP.

AC 057012.

DT 31-AUG-1994 (first entry)

DE Ptdins 3-kinase 110 KD catalytic subnt cDNA.

KW 110 KD catalytic subunit; phosphatidylinositol 3-kinase;

KW transformation; Schizosaccharomyces pombe; nmt promoter; thiamine;

KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;

KW blood vessel plaques; ss.

OS Bos taurus.

FT Key Location/Qualifiers

FT cds 1..3207

FT /*tag= a

FT /*product= p110

PN M09403609-A.

PD 17-FEB-1994.

PR 05-AUG-1993; G01651.

PR 05-AUG-1992; GB-016654.

PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.

PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;

DR WPI: 94-065697/08.

DR P-PSDB: R46294.

PT Eukaryotic cells transformed with mammalian phospholipid or

PT protein kinase DNA - useful in assays for compounds involved in

PT cell growth regulation and for treating cancers

PS Disclosure; Fig 1; 71pp; English.

CC This sequence encodes the 110 KD catalytic subunit of the phosphatidylin

CC inositol (Ptdins) 3-kinase. This sequence was transformed into

CC Schizosaccharomyces pombe cells under the regulatory control of the

CC nmt promoter in an embodiment of the invention. In the presence of

CC thiamine the promoter is inactive and the cells carrying the Ptdins

CC catalytic subunit plasmid grow as the parental strain. In the absence

CC of thiamine the nmt promoter functions and the Ptdins 3-kinase

CC catalytic subunit is induced. Ptdins activity is substantially

CC increased under these conditions. Cells containing constructs such as

CC this, are useful in assays for detecting compounds involved in cell

CC growth regulation. It is also used as the basis for detecting

CC compounds for treating cancers and the formation of blood vessel

CC plaques.

SO Sequence 3498 BP; 1148 A; 627 C; 721 G; 1002 T;

Query Match 50.0%; Score 19; DB 1; Length 3498;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 20 AGCGTTCTTTAGCCATCA 38

DB 878 AGCGTTCTTTAGCCATCA 860

RESULT 8

ID X13394/c

ID X13394 standard; DNA; 1332 BP.

```

AC      X13394; (first entry)
DE      Enterococcus faecalis genome config SEQ ID NO:457.
KW      Enterococcus faecalis; config; detection; Enterococcal infection;
OS      vaccine; attenuation; computer readable medium; ds.
PN      W09850555-A2.
PD      12-NOV-1998.
PR      04-MAY-1998; U08985.
PR      14-NOV-1997; U5-066009.
PR      06-MAY-1997; U5-044031.
PR      16-MAY-1997; U5-046555.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Dillon PJ, Kunsch CA;
PR      WPI; 99-045171/04.
PT      New isolated Enterococcus faecalis polynucleotides and polypeptides
PT      - used to develop products for the detection of Enterococcus and for
PT      use in vaccines for prevention or attenuation of Enterococcus
PT      infection.
PS      Claim 1; Page 1690-1691; 2084pp; English.
CC      A computer readable medium has been developed which has recorded on it
CC      982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC      X12938 to X13919 represent these nucleotide sequences which are primary
CC      nucleotide sequences, also known as configs. The computer-based system
CC      can identify fragments of the Enterococcus faecalis genome with
CC      commercial importance. The products can be used to detect the presence
CC      of Enterococcus faecalis in samples. They can also be used for
CC      diagnosing Enterococcal infection in an animal and monitoring
CC      progression of disease, and for identifying agents which can be used to
CC      modulate the growth or pathogenicity of Enterococcus faecalis, or
CC      another related organism, in vivo or in vitro. In particular the
CC      polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC      can be used in vaccines to prevent or attenuate an Enterococcal
CC      infection.
SQ      Sequence 1332 BP; 479 A; 190 C; 227 G; 431 T;

Query Match 49.5%; Score 18.8; DB 1; Length 1332;
Best Local Similarity 76.7%; Pred. No. 29;
Matches 23; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      9 GACATGCGCTAGGCTTCTTTAGGCATCA 38
DB      496 GTACATGCGCTAGTGTTTTACGATCA 467

RESULT 9
ID      N80161 standard; DNA; 1731 BP.
AC      N80161;
DE      19-OCT-1990 (first entry)
KW      Human papilloma virus 18 protein
KW      Human papilloma virus; HPV16; HPV18; malignant genital tumours; ss.
FH      Key Location/Qualifiers
FT      cds 99..572
FT      FT /*tag= a
FT      FT /label=HPV18 E6 protein
FT      cds 584..898
FT      FT /*tag= b
FT      FT /label=HPV 18 E7 protein
FT      cds 908..729
FT      FT /*tag= c
FT      FT /label=HPV 18 E1 protein
PN      DE3625257-A.
PD      04-FEB-1988.
PD      25-JUL-1986; 625257.
PR      23-JUL-1986; DE-624786,
PA      (BEHM) Behringwerke AG.
PI      Becker H, Meyer-Dulheuer K;
DR      WPI; 88-037095/06.
P-PSDB: P82076.
Expression prods. of human papilloma viruses type 16 and 18 -
used for the prodn. of diagnostics for HPV16 and HPV18 infections

```

```

PS      Disclosure: ; German.
CC      The expression prods. of HPV18, contained in this sequence, are used to
CC      produce antibodies to detect these. HPV16 expression prods are also used
CC      for antibody prodn. See also N80155-59.
SQ      Sequence 1731 BP; 610 A; 313 C; 403 G; 405 T;

Query Match
Best Local Similarity 48.9%; Score 18.6; DB 1; Length 1731;
Matches 21; Conservative 0; Pred. No. 38; Mismatches 4; Indels 0; Gaps 0

QY      9 GAACATGCGCTAGCGCTTCTTAGC 33
      |||||
      |||||

DB      1117 GAACATGCGCTGTGCTCTTAGC 1093

RESULT 10
T88772/c
ID      T88772 standard; DNA; 900 BP.
AC      T88772;
DR      23-MAR-1998 (first entry)
DE      Staphylococcus aureus ERA encoding DNA.
KM      GTP binding protein family; ERA; immune response; infection;
OS      diagnosis; antibody; respiratory tract; ss.
Staphylococcus aureus.
WO9730149-A1.
PD      21-AUG-1997.
PF      19-FEB-1997; U02547.
PR      20-FEB-1996; US-011888.
PA      (SMK ) SMITHKLINE BEECHAM CORP.
PI      Black MT, Burnham MKR, Hodgson JE, Knowles DJC,
PI      Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,
PI      Ward JM;
PI      WPI: 97-425028/39.
DR      P-PSDB; W31905.
PT      Staphylococcus aureus ERA gene - used to induce immune response
PT      against diseases related to S. aureus infection
PS      Claim 1; Fig 1; 46pp; English.
CC      The present sequence encodes a novel isolated GTP binding protein
CC      referred to as ERA. ERA can be used for diagnosing a disease related
CC      to ERA expression. The ERA protein can be used for inducing an
CC      immunological response in a mammal, to produce antibodies to protect
CC      the animal from disease. The disease is an infection by Staphylococcus
CC      aureus (or another bacteria), and include infections of the upper
CC      respiratory tract (e.g. otitis media, acute epiglottitis, thyroiditis),
CC      lower respiratory tract (e.g. lung abscess), cardiac (e.g. infective
CC      endocarditis), gastrointestinal tract (e.g. secretory diarrhoea, splenic
CC      abscess), central nervous system (e.g. cerebral abscess), eye (e.g.
CC      conjunctivitis), kidney and urinary tract (e.g. epididymitis, toxic
CC      shock syndrome), skin (e.g. folliculitis, cellulitis, wound infection),
CC      and bone and joint (e.g. septic arthritis).
SQ      Sequence 900 BP; 348 A; 118 C; 179 G; 255 T;

Query Match
Best Local Similarity 48.9%; Score 18.6; DB 1; Length 900;
Matches 24; Conservative 0; Pred. No. 33; Mismatches 9; Indels 0; Gaps 0

QY      4 GATCAGACATGCGCTAGCGCTTCTTAGCCAT 36
      |||
      |||
      |||

DB      684 GATATGACACGATCTGCGCTTCTTTAACCAT 652

RESULT 11
V21210/c
ID      V21210 standard; DNA; 58407 BP.
AC      V21210;
DR      10-NOV-1998 (first entry)
DE      Methanococcus jannaschii large circular extrachromosomal element.
KM      Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KM      genome; autotrophic; extrachromosomal element; identification; ds.
OS      Methanococcus jannaschii.
WO9807830-A2.
PN

```


PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057659.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DE, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,
PI MPI: 98-506364/43.
DR P-FSDB: W75006.
DR
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PS disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 519-521; 721pp: English.
CC This sequence represents a nucleic acid molecule designated Gene 155 from
CC the human CDNA clone H1XFJ55 (deposited as clone ATCC 97903 and ATCC
CC 209049) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W74731-W75006) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
CC Sequence 3276 BP; 811 A; 783 C; 928 G; 747 T;
CC

Query Match 48.9%; Score 18.6; DB 1; Length 3276;
Best Local Similarity 72.7%; Pred. No.44;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGATCAGACATGCTAGGCTTCTTACG 33
DB 1406 ATGATCTGACAGATCTCTAAGGCTCTGTACG 1374

RESULT 13
V59665/c
ID V59665 standard; DNA: 2933 BP.
AC V59665;
DT 19-JAN-1999 (first entry)
DE Human secreted protein gene 155 clone H1XFJ55.
KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW diagnosis; tissue; cancer; neurodegenerative disorder; leukaemia;
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW inflammation; ischaemic shock; Alzheimer's disease; osteoclast; AIDS;
KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW endocrine; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS Homo sapiens.
PN MO9839448-A2.
PD 11-SEP-1998.
PR 06-MAR-1998; U04493.
PR 02-OCT-1997; US-061060.
PR 07-MAR-1997; US-038621.
PR 07-MAR-1997; US-040161.

PR 07-MAR-1997; US-040162.
PR 07-MAR-1997; US-040163.
PR 07-MAR-1997; US-040333.
PR 07-MAR-1997; US-040334.
PR 07-MAR-1997; US-040336.
PR 07-MAR-1997; US-040626.
PR 11-APR-1997; US-043311.
PR 11-APR-1997; US-043312.
PR 11-APR-1997; US-043313.
PR 11-APR-1997; US-043314.
PR 11-APR-1997; US-043368.
PR 11-APR-1997; US-043369.
PR 11-APR-1997; US-043376.
PR 11-APR-1997; US-043378.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043659.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 13-JUN-1997; US-049610.
PR 08-JUL-1997; US-051826.
PR 16-JUL-1997; US-052874.
PR 18-AUG-1997; US-055724.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056633.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056845.
PR 22-AUG-1997; US-056862.
PR 22-AUG-1997; US-056864.
PR 22-AUG-1997; US-056872.
PR 22-AUG-1997; US-056874.
PR 22-AUG-1997; US-056875.
PR 22-AUG-1997; US-056876.

```

PR 22-AUG-1997; US-056877.
PR 22-AUG-1997; US-056878.
PR 22-AUG-1997; US-056879.
PR 22-AUG-1997; US-056880.
PR 22-AUG-1997; US-056881.
PR 22-AUG-1997; US-056882.
PR 22-AUG-1997; US-056883.
PR 22-AUG-1997; US-056884.
PR 22-AUG-1997; US-056885.
PR 22-AUG-1997; US-056886.
PR 22-AUG-1997; US-056887.
PR 22-AUG-1997; US-056888.
PR 22-AUG-1997; US-056889.
PR 22-AUG-1997; US-056890.
PR 22-AUG-1997; US-056891.
PR 22-AUG-1997; US-056892.
PR 22-AUG-1997; US-056893.
PR 22-AUG-1997; US-056894.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.
PR 22-AUG-1997; US-056909.
PR 22-AUG-1997; US-056910.
PR 22-AUG-1997; US-056911.
PR 05-SEP-1997; US-057650.
PR 05-SEP-1997; US-057659.
PR 05-SEP-1997; US-057761.
PR 12-SEP-1997; US-058785.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
PI Feng P, Ferrite AM, Fischer CL, Florence KA, Greene JM, Hu JS,
PI Kiyam H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
PI Ruben EM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,
DR WPI: 98-506364/43.
P-PSDB: W74883.
PT New isolated human genes and the secreted polypeptide(s) they encode
PT - useful for diagnosis and treatment of e.g. cancers, neurological
PT disorders, immune diseases, inflammation or blood disorders
PS Claim 1: Page 391-393; 721pp; English.
CC This sequence represents a nucleic acid molecule designated gene 155 from
CC the human cDNA clone HMXF155 (deposited as clone ATCC 97903 and ATCC
CC 209049) which encodes a secreted human protein. The gene can be used to
CC generate fusion proteins by linking to the gene to a human immunoglobulin
CC Fc portion (e.g. V59502) for increasing the stability of the fused
CC protein as compared to the human protein only.
CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: V59511-V59812; amino acid sequences W4731-W75026) which
CC are useful for preventing, treating or ameliorating medical conditions
CC e.g. by protein or gene therapy. Also, pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new polynucleotides.
CC Specific uses are described for each of the 186 polynucleotides, based on
CC which tissues they are most highly expressed in (see V59511 for described
CC uses).
SQ Sequence 2933 BP; 726 A; 715 C; 833 G; 651 T;

Query Match 48.9%; Score 18.6; DB 1; Length 2933;
Best Local Similarity 72.7%; Pred. No. 43;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGATCAGAACATGCTAGGCTTCTTACG 33
DB 1409 ATGATCTGAGAGATCTTAAGCGCTGTAGC 1377

```

```

FT misc-feature 1561..1620
FT /tag- a
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 3361..3420
FT /tag- b
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
FT misc-feature 5161..5220
FT /tag- c
FT /note- "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA,
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S aureus vaccines
PS Claim 1: Page 1297-1300; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 6234 BP; 2215 A; 822 C; 1171 G; 1852 T;

Query Match 48.9%; Score 18.6; DB 1; Length 6254;
Best Local Similarity 72.7%; Pred. No. 50;
Matches 24; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 GATCAGAACATGCTAGGCTTCTTACCAT 36
DB 5938 GATATGACACAGATCTTGCCTTCTTACCAT 5906

```

```

RESULT 14
V74715/c
ID V74715 standard; DNA: 6254 BP.
AC V74715;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #404.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers

```

```

RESULT 15
N90338
ID N90338 standard; cDNA: 12923 BP.
AC N90338;
DT 29-MAR-1992 (first entry)
DE Sequence of human muscular dystrophy (MD) cDNA.
DE Dystrophin; muscular dystrophy; probe; antibody; diagnosis;
KW prenatal; heterozygote; gene therapy; genetic screening;
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 209..12923
FT /*tag- a

```

PN W08906286-A.
PD 13-JUL-1989.
PF 16-DEC-1988; U04504.
PR 22-DEC-1987; US-136618.
PA (CHIL-) CHILDRENS MED CENT.
PI Kunkel LM, Monaco A, Hoffman EP, Koenig M;
DR WPI; 89-220587/30.
P-PSDB; P90373.
PT Muscular dystrophy gene - used for prep. of probes, dystrophin
PT polypeptide and antibodies for diagnosis and therapy of muscular
PT dystrophy
PS Disclosure: Fig 5: 68pp: English.
CC The inventors claim an MD probe comprising a purified ss NA, SQ which
CC hybridises to at least a part of the MD gene; pure dystrophin (DS)
CC polypeptide, purified NA encoding DS and antibodies (Ab) to DS. The
CC probes are equal to or greater than 10b of one of 12 cDNA sequences
CC deposited as ATCC 58666-57677. The MD gene is human, or a murine Dmd
CC gene.
SQ Sequence 12923 BP; 4296 A; 2613 C; 2963 G; 3044 T;

Query Match 48.4%; Score 18.4; DB 1; Length 12923;
Best Local Similarity 69.4%; Pred. No. 72;
Matches 25; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 2 AGGATCAGAACATGCTTAGGCTTCTTTAGCCATC 37
DB 2688 AGTATCAGAACATCATCATGCTTCTATATATCAGC 2723

Search completed: September 13, 1999, 16:22:05
Job time: 4025 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:13:58 ; Search time 849.94 Seconds

(without alignments)
88.190 Million cell updates/sec

Title: US-09-325-095-2425.SEQ

Perfect score: 38
Sequence: 1 AAGATCAGACAAATGCCCTAGGCTTCTTAGCCARCA 38

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	21.8	57.4	546	54	HSW010244	AL045394 Homo sapi
C 2	21	55.3	249	37	AA716393	AA716393 z662d06.s
C 3	21	55.3	355	39	AA903710	AA903710 CR60a11.s
C 4	21	55.3	279	44	A1282767	A1282767 q184D06.x
C 5	20.8	54.7	325	46	A1029617	A1029617 U1-R-C0-1
C 6	20.6	54.2	371	21	R02605	R02605 yef6a06.r1
C 7	20.6	54.2	411	33	AA446728	AA446728 zw84f03.r
C 8	20.6	54.2	484	40	AA922889	AA922889 o151d12.s
C 9	20.4	53.7	700	54	HSW010166	AL045316 Homo sapi
C 10	20.2	53.2	717	22	R81085	R81085 EST22e01 WA
C 11	20.2	53.2	426	24	H84400	H84400 Y101a10.r1
C 12	20.2	53.2	400	27	AA018084	AA018084 z651a10.s
C 13	20.2	52.6	390	29	AA172139	AA172139 z2d4e05.r
C 14	20	52.6	307	46	A1425878	A1425878 mg09h08.x
C 15	19.8	52.1	628	38	AA799278	AA799278 EST18875
C 16	19.8	52.1	1006	48	A1573683	A1573683 U164e07.Y
C 17	19.6	51.6	581	25	N58217	N58217 yv66h06.s1
C 18	19.6	51.6	416	25	N83070	N83070 TGEST2y60c1
C 19	19.6	51.6	142	26	W70126	W70126 z668c07.s1
C 20	19.6	51.6	205	29	AA193522	AA193522 zr41h09.s
C 21	19.6	51.6	542	30	AA214230	AA214230 zq90f07.r
C 22	19.6	51.6	330	33	AA426031	AA426031 zv52d07.s
C 23	19.6	51.6	378	37	AA719187	AA719187 z134b02.s
C 24	19.6	51.6	589	38	AA264378	AA264378 LD08008.s
C 25	19.6	51.6	546	38	AA438766	AA438766 LD13119.s
C 26	19.6	51.6	790	39	AA843153	AA843153 AK06d03.s
C 27	19.6	51.6	439	42	A1129164	A1129164 OX87h05.s
C 28	19.6	51.6	474	43	A1183972	A1183972 q669f06.x
C 29	19.6	51.6	429	47	A1512524	A1512524 LD44422.s
C 30	19.6	51.6	458	49	A1633740	A1633740 t128a04.x
C 31	19.6	51.6	451	50	AU065351	AU065351 AU065351
C 32	19.4	51.1	335	20	T30439	T30439 EST16732 Hu
C 33	19.4	51.1	325	21	R05870	R05870 yeg3b08.r1
C 34	19.4	51.1	629	29	AA141078	AA141078 CR01174.3
C 35	19.4	51.1	1019	35	C23297	C23297 C23297 Japa
C 36	19.4	51.1	526	38	AA786092	AA786092 j5c09a1.r
C 37	19.4	51.1	646	39	AA820961	AA820961 LD24919.s
C 38	19.4	51.1	586	40	AA948780	AA948780 LD27409.s
C 39	19.4	51.1	803	41	A1066025	A1066025 mgae0005a
C 40	19.4	51.1	423	43	A1209388	A1209388 q0d03a1.r
C 41	19.4	51.1	328	45	A1331461	A1331461 fag3b03.y
C 42	19.4	51.1	439	45	A1332069	A1332069 fag3b03.x
C 43	19.4	51.1	721	45	AU004032	AU004032 AU004032
C 44	19.4	51.1	534	46	A155453	A155453 LD24340.3
C 45	19.4	51.1	276	47	A1540728	A1540728 t190a03.x

ALIGNMENTS

RESULT 1
ID HSW010244/c standard; RNA; EST; 546 BP.
XX
AC AL045394;
XX
SV
XX AL045394.1

NI e1403868
 XX 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX Homo sapiens mRNA; EST DKFZp434D205_r1 (from clone
 DE DKFZp434D205)
 XX EST: expressed sequence tag.
 XX Homo sapiens (human)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 CC Primates; Catarrhini; Homidae; Homo.
 XX [1]
 RN 1-546
 RP Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
 RA MRS, Am Klopferstritz 18a D-82152 Martinsried, GERMANY
 RL
 CC Clone from S. Wiemann, sequenced by Qiagen within the CDNA
 CC sequencing consortium of the German Genome Project
 CC s1 sequence also available
 CC This clone is available at the RZPD in Berlin
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
 XX
 XX Key Location/Qualifiers
 FH
 FT source 1..546
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="DKFZp434D205"
 FT /clone_lib="434 (synonym: htes3). Vector pSport1; host
 FT DH10B; sites NotI + SalI"
 FT /dev_stage="adult"
 FT /tissue_type="testis"
 FT
 XX Sequence 546 bp; 156 A; 123 C; 143 G; 124 T; 0 other;
 SO

Query Match 57.4%; Score 21.8; DB 54; Length 546;
 Best Local Similarity 78.8%; Pred. No. 18;
 Matches 26; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 5 ATCAGACACATGCTAGGCTTCTTACCATC 37
 Db 118 ATCAGACACATCTGCGCTTCTCAGACATC 86

RESULT 2
 AA716393 249 bp mRNA EST 29-DEC-1997
 LOCUS 2962d06.s1 Soares_fetal_heart.NBH19W Homo sapiens cDNA clone
 DEFINITION IMAGE:397931.3', mRNA sequence.
 ACCESSION AA716393
 MID 92728667
 VERSION AA716393.1 GI:2728667
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 249)
 AUTHORS Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
 Martin,J., Moore,B., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
 Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
 WashU-NCI human EST Project
 JOURNAL Unpublished (1997)
 COMMENT On May 8, 1995 this sequence version replaced gi:801246.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: east@wustl.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 135.
 FEATURES
 source
 1..249
 /organism="Homo sapiens"
 /db_xref="GDB:1305754"
 /db_xref="taxon:9606"
 /clone="IMAGE:397931"
 /clone_lib="Soares_fetal_heart_NBH19W"
 /sex="unknown"
 /dev_stage="19 weeks"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: heart; Vector: pT73D (Pharmacia) with a
 modified polylinker; site_1: Not I; site_2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTACCAATCTGAGAGGAGGCGGCGCATCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adapters (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pT73 vector
 (Pharmacia). Library went through one round of
 normalization to a Cot = 5. Library constructed by
 M. Fatima Bonaldo. This library was constructed from the
 same fetus as the fetal lung library; Soares fetal lung
 NBH19W."

BASE COUNT 68 a 62 c 47 g 72 t
 ORIGIN

Query Match 55.3%; Score 21; DB 37; Length 249;
 Best Local Similarity 73.0%; Pred. No. 37;
 Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 2 AGATCAGACACATGCTAGGCTTCTTACCATCA 38
 Db 8 AGATCTTACCATGCGCTAGGCTTCTTACCATTA 44

RESULT 3
 AA903710 355 bp mRNA EST 08-APR-1998
 LOCUS 0606a11.s1 NCI-CGAP-GC4 Homo sapiens cDNA clone IMAGE:1518132.3'
 DEFINITION similar to contains Alu repetitive element; mRNA sequence.
 ACCESSION AA903710
 MID 93038833
 VERSION AA903710.1 GI:3038833
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 355)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1407050.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/URL at: www-bio.lnl.gov/bdrip/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 262.

FEATURES
Source
Location/Qualifiers

1..335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1518332"
/clone_id="NCI-CGAP-GC4"
/tissue="type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"

BASE COUNT
ORIGIN
98 a 83 c 76 g 98 t

Query Match 55.3%; Score 21; DB 39; Length 355;
Best Local Similarity 73.0%; Pred. No. 38;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 2 AGGATCAGACATGCTAGGCTTCTTTAGCCATCA 38
||||| ||||||| ||| ||| |||
Db 7 AGGATCTTACCATGCTAGGCTTCTTTAGCCATCA 43

RESULT 4
LOCUS A1282767 279 bp mRNA EST 21-DEC-1998
DEFINITION q184b06.x1 NCI-CGAP.Co14 Homo sapiens CDNA clone IMAGE:1961939 3'
similar to contains Alu repetitive element; contains element LTR6
repetitive element ; mRNA sequence.
A1282767
NID 93821000
VERSION A1282767.1 GI:3921000
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
1 (bases 1 to 279)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On May 8, 1995 this sequence version replaced gi:800256.

JOURNAL
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nlh.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/URL at: www-bio.lnl.gov/bdrip/image/image.html

FEATURES
Source
Insert Length: 1300 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 237.
Location/Qualifiers
1..279

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1961939"
/clone_id="NCI-CGAP-Co14"
/tissue="type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: PCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

BASE COUNT
ORIGIN
74 a 71 c 55 g 79 t

Query Match 55.3%; Score 21; DB 44; Length 279;
Best Local Similarity 73.0%; Pred. No. 37;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

OY 2 AGGATCAGACATGCTAGGCTTCTTTAGCCATCA 38
||||| ||||||| ||| ||| |||
Db 7 AGGATCTTACCATGCTAGGCTTCTTTAGCCATCA 43

RESULT 5
LOCUS A1029617/c 325 bp mRNA EST 09-MAR-1999
DEFINITION UI-R-C0-1Y-e-09-0-UI.s1 UI-R-C0 Rattus norvegicus CDNA clone
UI-R-C0-1Y-e-09-0-UI 3', mRNA sequence.
A1029617
ACCESSION 94299791
VERSION A1029617.1 GI:4299791
KEYWORDS EST.
SOURCE Norway rat.
ORGANISM Rattus norvegicus

REFERENCE
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 325)
Bonaldo,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
JOURNAL
MEDLINE On Jun 22, 1998 this sequence version replaced gi:3247443.

COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: mesares@blue.weeg.uiowa.edu

The sequence tag present in the CDNA between the NotI site and the
oligo-dT track served to identify it as a clone from the normalized
adult 12-Day-Embryo library. CDNA Library Preparation: M. Fatima
Bonaldo, Ph.D. Clone distribution: clones will be available through
Research Genetics
Seq primer: M13 Forward.

FEATURES
Source
Location/Qualifiers

1..325
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-C0-1Y-e-09-0-UI"
/clone_id="UI-R-C0"
/der_stage="adult"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pRT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C0
library is a subtracted library derived from the UI-R-A1
and UI-R-E1 libraries. The UI-R-A1 library consisted of a
mixture of individually tagged normalized libraries
constructed from rat placenta, adult lung, brain, liver,
kidney, heart, spleen, ovary, and muscle. The UI-R-E1

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Palma Bonaldo. "

BASE COUNT 115 a 57 c 63 g 176 t
 ORIGIN

Query Match 54.2%; Score 20.6; DB 33; Length 411;
 Best Local Similarity 74.3%; Pred. No. 56;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 2 AGGATCAGACAAATGCTTACGCTTTTACCAT 36
 Db 328 AAGAGAGACAAATATATAGTTCTTATGAT 362

RESULT 8
 LOCUS AA922889 484 bp mRNA EST 19-MAY-1998
 DEFINITION O151912.s1 NCI-CGAP_HN3 Homo sapiens cDNA clone IMAGE:1486246 3' similar to SW:EC01_HUMAN O13011 PROBABLE PROXISOMAL ENOYL-COA HYDRATASE ;, mRNA sequence.

ACCESSION AA922889
 NID 93070198
 VERSION AA922889.1 GI:3070198
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NCI/NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute / National Institute of Dental Research, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 Unpublished (1997)
 On Oct 20, 1997 this sequence version replaced gi:2520646.

COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: John Ensley, M.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILM at: www.bio.lnl.gov/bbrp/image/image.html

Insert Length: 1258 Std Error: 0.00
 Seq primer: -40m13 fwd. Et from Amersham
 High quality sequence stop: 260.

FEATURES
 source
 1..484
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="IMAGE:1486246"
 /clone_lib="NCI-CGAP_HN3"
 /tissue_type="squamous cell carcinoma from base of tongue"
 /lab_host="SOBR (kanamycin resistant)"
 /note="Organ: tongue; Vector: Bluescript SK-; Site: 1; EcorI: Site 2: XhoI; Cloned unidirectionally. Primer: O150 dt. Average insert size 1.0 kb. 5' adaptor sequence: 5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5' (GA)10ACTACTCGAGCTTTTCTTTTCTTTT 3' "

BASE COUNT 99 a 134 c 137 g 114 t
 ORIGIN

Query Match 54.2%; Score 20.6; DB 40; Length 484;
 Best Local Similarity 74.3%; Pred. No. 57;
 Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 GATCAGAACAAATGCTTACGCTTTTACCATCA 38
 Db 446 GACCAGCCCAATGCCAGGCGCTCTGACGCATCA 480

RESULT 9
 HSM010166/c standard; RNA; EST; 700 BP.

AC AL045316;
 SV AL045316.1
 NI e1403790

DT 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)

DE Homo sapiens mRNA; EST DKFZP434A015_r1 (from clone DKFZP434A015)

XX EST; expressed sequence tag.

OS Homo sapiens (human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

RA Dueterhoeft A., Lauber J., Mewes W., Gassenhuber J., Wiemann S.;
 RP 1-700
 [1]
 Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.

RL M195, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
 RL CC
 CC Clone from S. Wiemann, sequenced by Qiagen within the CDNA sequencing consortium of the German Genome Project

CC This clone is available at the RZPD in Berlin
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
 CC XX
 FH Key Location/Qualifiers

FT source 1..700
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="DKFZP434A015"
 FT /clone_lib="434 (synonym: htes3). Vector pSport1; host DH10B; sites NotI + SalI"
 FT /dev_stage="adult"
 FT /tissue_type="testis"

SO Sequence 700 BP; 164 A; 156 C; 125 G; 254 T; 1 other;

Query Match 53.7%; Score 20.4; DB 54; Length 700;
 Best Local Similarity 80.0%; Pred. No. 72;
 Matches 24; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGATCAGAACAAATGCTTACGCTTTCTT 30
 Db 523 AAGATTAAGACCAAGCCGAGCTTTTGT 494

RESULT 10

LOCUS R81085 717 bp mRNA EST 12-JUN-1996
 DEFINITION EST222601 WATM1 Homo sapiens cDNA clone 222601, mRNA sequence.

ACCESSION R81085
 NID 9857688
 VERSION R81085.1 GI:857688

KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 717)
 Bouillaud, F.
 Study of expressed sequence tags in adipose tissue 1995
 Unpublished (1995)
 On May 8, 1995 this sequence version replaced gi:801224.

TITLE JOURNAL
 COMMENT

CONTACT: Frederic Bouillaud
 Centre de Recherche sur l'Endocrinologie moléculaire et le
 développement
 CNRS
 9, Rue Jules Hetzel, Meudon Bellevue, 92190 France
 Tel: 33 1 45 07 52 87
 Fax: 33 1 45 07 58 90
 Email: bouillaud@infobio.fr
 automatic cycle sequencing of PCR amplified insert
 Insert length: 851 Std Error: 0.00
 High quality sequence stop: 290.

FEATURES
 source
 1. 717
 /organism="Homo sapiens"
 /strain="caucasian"
 /db_xref="taxon:9606"
 /clone="22e01"
 /note="Vector: lambda g11; Site 1: EcoRI; White adipose
 tissue, sub cutaneous, adult, female. Purification of
 polyA mRNA, first strand priming with random
 oligonucleotides. Ligation of adaptors EcoRI/NotI, cloning
 in EcoRI site of lambda g11."

BASE COUNT 196 a 172 c 192 g 157 t
 ORIGIN

Query Match 53.2%; Score 20.2; DB 22; Length 717;
 Best Local Similarity 88.0%; Pred. No. 87;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 14 ATGCTAGGCTTCTTACCATCA 38
 ||||| ||||| ||||| ||||| ||
 Db 258 ATGCCAGGCTTCATTAGCCACCA 234

RESULT 11
 H84400 426 bp mRNA EST 13-NOV-1995
 LOCUS Y101a10.r1 Soares retina N2B5HR Homo sapiens cDNA clone
 DEFINITION IMAGE:223002 5' similar to gb:D90209 DNA-BINDING PROTEIN TAXREB67
 (HUMAN); mRNA sequence.

ACCESSION H84400
 NID 91063071
 VERSION H84400.1 GI:1063071
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 426)
 Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
 Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
 Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
 Trevasaks, E., Waterston, R., Williamson, A., Wohlmann, P., and
 Wilson, R.
 The WashU-Merck EST Project
 Unpublished (1995)
 On May 18, 1995 this sequence version replaced gi:811191.

TITLE JOURNAL
 COMMENT

CONTACT: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

Fax: 314 286 1810
 Email: est@wustl.edu
 High quality sequence stops: 260
 Source: IMAGE Consortium, LNL
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert Length: 1659 Std Error: 0.00
 Seq primer: M13Rpi
 High quality sequence stop: 260.

FEATURES
 source
 1. 426
 /organism="Homo sapiens"
 /db_xref="GDB:3854036"
 /db_xref="taxon:9606"
 /clone="IMAGE:223002"
 /clone_1ib="Soares retina N2B5HR"
 /sex="male"
 /tissue="retina"
 /dev_stage="55 year old"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: eye; Vector: pTZ19 (Pharmacia) with a
 modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dT) primer [5'
 TGTTCACATCGAGTGGAGCGCGCGCGCTTTTCTTTTCTTTT 3']
 double-stranded cDNA was size selected, ligated to Eco RI
 adaptors (Pharmacia), digested with Not I and cloned into
 the Not I and Eco RI sites of a modified pTZ19 vector
 (Pharmacia). The retinas were obtained from a 55 year old
 caucasian and total cellular poly(A)+ RNA was extracted 6
 hrs after their removal. The retina RNA was kindly
 provided by Roderick R. McInnes M.D. Ph.D. from the
 University of Toronto. Library constructed by Bento
 Soares and M. Fatima Bonaldo."

BASE COUNT 89 a 117 c 113 g 99 t 8 others
 ORIGIN

Query Match 53.2%; Score 20.2; DB 24; Length 426;
 Best Local Similarity 73.5%; Pred. No. 83;
 Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGGATCAGACATGCTTCTTACCC 34
 ||||| ||||| ||||| ||||| ||
 Db 371 AATCATCTTAGAGACCTGCTTCTTACCC 338

RESULT 12
 AA018084 400 bp mRNA EST 30-JAN-1997
 LOCUS zef1a10.s1 Soares retina N2B4HR Homo sapiens cDNA clone
 DEFINITION IMAGE:362490 3' mRNA sequence.

ACCESSION AA018084
 NID 91481339
 VERSION AA018084.1 GI:1481339
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 400)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favell, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Plenge, C., Rifkin, L.,
 Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Tillery-Wieg, J.,
 Trevasaks, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

TITLE JOURNAL
 MEDLINE
 COMMENT

On Apr 14, 1993 this sequence version replaced gi:785894.

CONTACT: Wilson RK

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 2291 Std Error: 0.00
Seq primer: -40M13 fwd. from AmerSham
High quality sequence stop: 310.

FEATURES

source

1.400

Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:1279193"
/db_xref="taxon:9606"
/clone="IMAGE:362490"
/clone_1bp="Soares retina N2b4HR"
/sex="male"
/tissue_type="retina"
/dev_stage="55 year old"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: eye; Vector: pRT73D (Pharmacia) with a
modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCATCTGAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pRT73 vector
(Pharmacia). The retinas were obtained from a 55 year old
Caucasian and total cellular poly(A)+ RNA was extracted 6
hrs after their removal. The retina RNA was kindly
provided by Roderick R. McInnes M.D. Ph.D. from the
University of Toronto. Library constructed by Bento
Soares and M. Fatima Bonaldo."

BASE COUNT 113 a 75 c 84 g 126 t 2 others

ORIGIN

Query Match 53.2%; Score 20.2; DB 27; Length 400;
Best Local Similarity 73.5%; Pred. No. 83;
Matches 25; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 AAGGATCAGAACATGCTAGCGTTCTTATGCC 34

Db 391 AATGGTAGAAATGCCANGCCTTCTTATCC 358

RESULT 13

AA172139

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

390 bp mRNA EST 23-DEC-1996
zpz4e05.x1 StrataGene neuroepithelium (#937231) Homo sapiens CDNA
clone IMAGE:610400 5', mRNA sequence.
AA172139
51751198
AA172139.1 GI:1751198
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 390)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chippelli, B.,
Chissole, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On May 18, 1995 this sequence version replaced gi:811357.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -28M13 rev2 from AmerSham
High quality sequence stop: 352.

FEATURES

source

1.390

Location/Qualifiers

/organism="Homo sapiens"
/db_xref="GDB:4625659"
/db_xref="taxon:9606"
/clone="IMAGE:610400"
/clone_1bp="Stratagene neuroepithelium (#937231)"
/dev_stage="Ntera-2/RA neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site: 1: EcoRI; Site: 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT. NT2
cells (Ntera-2/cl.D1) induced with Retinoic Acid for 24
hours. Average insert size: 1.5 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GATTTCGCGACG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 143 a 67 c 54 g 126 t

ORIGIN

Query Match 52.6%; Score 20; DB 29; Length 390;
Best Local Similarity 72.2%; Pred. No. 1e+02;
Matches 26; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 3 GGATCAGAACATGCTAGCGTTCTTATGCCATCA 38

Db 242 GCATCAGAACATGCTAGCGTTCTTATGCCATCA 277

RESULT 14

A1425878/c

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

307 bp mRNA EST 09-MAR-1999
mg09h08.x1 Soares mouse embryo NDMEL3.5 14.5 Mus musculus CDNA
clone IMAGE:423327 3', mRNA sequence.
A1425878
94271809
A1425878.1 GI:4271809
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 307)
Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Petersen, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
On May 7, 1998 this sequence version replaced gi:3118818.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end and
High quality sequence stop: 273.
Location/Qualifiers

```

source
1.307
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="Scares mouse embryo NbME13.5.14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer 15'
TGTTCACATCTGAAGTGGAGCGCGCGGAAATTTTCTTTTCTTTT
T 3', on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."

BASE COUNT      111 a      49 c      46 g      101 t
ORIGIN

Query Match      52.6%; Score 20; DB 46; Length 307;
Best Local Similarity 82.1%; Pred. No. 98;
Matches 23; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy      3 GGATCAGACAAATGCTAGGCTTCTTT 30
1 ||||| ||||| ||||| ||||| |||||
Db      159 GAATCAGAAAATGCCAGCCTTTT 132

RESULT 15
AA799278      628 bp      mRNA      EST      30-APR-1998
LOCUS      AA799278/c
DEFINITION      EST188775 Normalized rat heart, Bento Soares Rattus sp. cDNA clone
RHEA05 5' end, mRNA sequence.
ACCESSION      AA799278
NID      92862233
VERSION      AA799278.1 GI:2862233
KEYWORDS
SOURCE
ORGANISM      Rattus sp.
Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 628)
Lee,N.H., Glodet,A., Chandra,I., Mason,T.M., Quackenbush,J.,
Kerlavage,A.R. and Adams,W.D.
Rat Genome Project: Generation of a Rat EST (RESE) Catalog & Rat
Gene Index
Unpublished (1998)
JOURNAL
COMMENT      On Jan 19, 1998 this sequence version replaced gi:2286418.
*Other-ESTs: EST188774
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3529
Fax: (301)-838-0208
Email: nhlee@tigr.org
Seq primer: M13 Reverse.
FEATURES
source
1.628
Location/Qualifiers
/organism="Rattus sp."
/db_xref="ATCC (1host):2006305"
/db_xref="taxon:10118"
/clone_1lb="RHEA05"
/clone_1lb="Normalized rat heart, Bento Soares"
/note="Organ: heart; Vector: pT73pac; Site_1: EcoRI;
Site_2: NotI"

```

```

BASE COUNT      169 a      163 c      155 g      140 t      1 others
ORIGIN

Query Match      52.1%; Score 19.8; DB 38; Length 628;
Best Local Similarity 77.4%; Pred. No. 1.3e+02;
Matches 24; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Oy      1 AAGATCAGACAAATGCTAGGCTTCTTTA 31
1 ||||| ||||| ||||| ||||| |||||
Db      493 AGGATTAGACAAATCCCTCCTTCTTTA 463

Search completed: September 13, 1999, 16:14:01
Job time: 5460 sec

```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:55:59 ; Search time 539.84 Seconds

(without alignments)
135.498 Million cell updates/sec

Title: US-09-325-095-29

Perfect score: 23

Sequence: 1 RAAATGCCRAATCDATRTGMA 23

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sy.*
15: gb_un.*
16: gb_un.*
17: gb_y1.*
18: em_fun.*
19: em_hlg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_st.*
31: em_sy.*
32: em_un.*
33: em_v1.*
34: gb_hlg1.*
35: gb_hlg2.*
36: gb_in1.*
37: gb_in2.*
38: em_ba1.*
39: em_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	18.2	79.1	111597	11	HSU85196	Homo sapien

C	2	18.2	79.1	247877	11	HUA0000661	AE0000661 Homo sapi
C	3	18.2	79.1	42467	37	CEL26F12	U55373 Caenorhabd1
C	4	18.2	79.1	37675	37	CELM01G5	AF078786 Caenorhab
C	5	17.4	75.7	297070	34	CEY47H10	Z95311 Caenorhabd1
C	6	17.4	75.7	21816	36	CEC47F8	AL009246 Caenorhab
C	7	17.2	74.8	155147	9	HS992D9	AL008932 Human DNA
C	8	17.2	74.8	218336	11	AF067844	AF067844 Homo sapi
C	9	17	73.9	97015	1	BSGENR	X73124 B. subtilis
C	10	17	73.9	212150	1	BSUB0020	Z99123 Bacillus su
C	11	17	73.9	3650	7	YSICMET10A	L26504 Saccharomyc
C	12	17	73.9	126839	10	HS19408	AL031769 Human DNA
C	13	17	73.9	7215	36	DMP13K68D	X92892 D. melanog
C	14	17	73.9	10213	36	PFU27338	U27338 Plasmodium
C	15	17	73.9	43584	37	CELB0207	U97196 Caenorhabd1
C	16	17	73.9	3606	37	DMU52192	U52192 Drosophila
C	17	16.6	72.2	3606	4	AB008857	AB008857 Fugu rubr
C	18	16.6	72.2	38	5	A37242	A37242 Sequence 11
C	19	16.6	72.2	87841	7	AB005247	AB005247 Arabidops
C	20	16.6	72.2	85791	7	AB016888	AB016888 Arabidops
C	21	16.6	72.2	69748	7	SCRRAC11	X78993 S. cerevisia
C	22	16.6	72.2	1600	7	SCYBR107C	Z35976 S. cerevisia
C	23	16.6	72.2	3390	7	SCYBR108W	Z35977 S. cerevisia
C	24	16.6	72.2	2224	7	SCYOR033W	Z74775 S. cerevisia
C	25	16.6	72.2	25380	7	SPAC1P8	Z81312 S. pombe chr
C	26	16.6	72.2	2043	7	YSCME1G	L39015 Saccharomyc
C	27	16.6	72.2	3485	8	AF025434	AF025434 Papaver s
C	28	16.6	72.2	126337	8	ATAC006067	AC006067 Arabidops
C	29	16.6	72.2	85444	9	AP000023	AP000023 Homo sapi
C	30	16.6	72.2	153147	9	HS13D10	AL021407 Homo sapi
C	31	16.6	72.2	126474	9	HS181N1	Z82899 Human DNA s
C	32	16.6	72.2	130705	9	HS232L22	Z73986 Human DNA s
C	33	16.6	72.2	113704	9	HS267P19	Z75889 Human DNA s
C	34	16.6	72.2	132229	9	HS447N6	Z82211 Human DNA s
C	35	16.6	72.2	31523	9	HS656F14	AL034416 Human DNA
C	36	16.6	72.2	123004	10	HS196E23	Z97632 Human DNA s
C	37	16.6	72.2	95855	10	HSAC000115	AC000115 Human BAC
C	38	16.6	72.2	39876	10	HSU83C4	Z70050 Human DNA s
C	39	16.6	72.2	41052	11	AC002499	AC002499 Human Cos
C	40	16.6	72.2	41876	11	AC004190	AC004190 Homo sapi
C	41	16.6	72.2	746	12	AF084643	AF084643 Mus muscu
C	42	16.6	72.2	1460	17	CAEVTTRNA	X64109 Caprine Art
C	43	16.6	72.2	1999	17	OLYNEWYORK	L19198 Ovine lenti
C	44	16.6	72.2	144247	34	AC005482	AC005482 Homo sapi
C	45	16.6	72.2	144247	34	AC005482	AC005482 Homo sapi

ALIGNMENTS

RESULT	1	HSU85196	111597 bp	DNA	PRI	27-MAY-1997
LOCUS		HSU85196	Homo sapiens BAC378, complete sequence.			
DEFINITION		U85196				
ACCESSION		U85196				
NID		92121230				
VERSION		U85196.1	GI:2121230			
KEYWORDS		HTG.				
SOURCE		human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS		Boysen,C., Simon,M.I. and Hood,L.				
TITLE		Analysis of the 1.1-kb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones				
JOURNAL		Genome Res. 7 (4), 330-338 (1997)				
MEDLINE		97264339				
REFERENCE		2 (bases 1 to 111597)				
AUTHORS		Boysen,C., Simon,M.I. and Hood,L.				
TITLE		Direct Submission				
JOURNAL		Submitted (14-JAN-1997) Molecular Biotechnology, University of Washington, Box 357730, Seattle, WA 98195, USA				
FEATURES		Location/Qualifiers				


```

gene      24820..25451
/gene="TCRAV38S1"
join(24820..24868,25112..>25412)
CDS       /gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="Old name TCRAV14S2"
/codon_start=1
/product="TCRAV38S1"
/protein_id="AAB69035.1"
/db_xref="PID:g2358061"
/db_xref="GI:2358061"
/translation="MTRVSLIMAVVSTCLESGMAQVVTOSOPMSVQEAETVLTST
YDTSSENYLFWYKQPPSRQMLIVIRQEAATKQDNATNRSVNFQRAKSFSLKTSIS
QLCDTMYECAPK"
misc_recomb 25413..25419
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="heptamer"
25420..25442
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="spacer"
25443..25451
/gene="TCRAV38S1"
/standard_name="TCRAV38S1"
/note="nonamer"
repeat_region complement(27096..27964)
/rpt_family="L1PB3"
complement(27821..29134)
/rpt_family="L1"
complement(30712..31212)
/rpt_family="MLT2CA"
32458..32606
/rpt_family="AluDb"
33599..33609
/rpt_type=tandem
/rpt_unit=AG
join(<34000..34048,34298..>34597)
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="Old name TCRAV14S1-hDV8S1"
/product="hADV38S2"
34000..34636
/gene="hADV38S2"
join(34000..34048,34298..>34597)
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="Old name TCRAV14S1-hDV8S1"
/codon_start=1
/product="hADV38S2"
/protein_id="AAB69036.1"
/db_xref="PID:g2358062"
/db_xref="GI:2358062"
/translation="MACGFLMALVISTCLEFSMAQVVTOSOPMSVQEAETVLTST
YDTSSENYLFWYKQPPSRQMLIVIRQEAATKQDNATNRSVNFQRAKSFSLKTSIS
QLCDTMYECAPK"
V_segment join(34000..34048,34298..>34597)
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="Old name TCRAV14S1-hDV8S1"
34598..34604
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="heptamer"
34605..34627
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="spacer"
34628..34636
/gene="hADV38S2"
/standard_name="hADV38S2"
/note="nonamer"
36263..36614
repeat_region

```

```

repeat_region /rpt_family="THEIA"
complement(36621..36703)
/rpt_family="MIR2"

Query Match 79.1%; Score 18.2; DB 11; Length 247877;
Best Local Similarity 65.2%; Pred. No. 1.1e+02;
Matches 15; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 BAARTGCCCAARTCDATRTGAA 23
Db 28569 AAAATGCTAAATCAATATGAA 28547
:::|::|::|::|::|::|::|::|
:::|::|::|::|::|::|::|::|

RESULT 3
CELF26F12 42467 bp DNA INV 02-JUN-1999
LOCUS Caenorhabditis elegans cosmid f26f12.
DEFINITION U55373
ACCESSION U55373
NID g3294506
VERSION U55373.1 GI:3294506
KEYWORDS
SOURCE
ORGANISM
Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 42467)
The C. elegans Genome Sequencing Consortium, Washington University
Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
Hinxton, U.K./C.
Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium
Science 282 (5396), 2012-2018 (1998)
99069613
Erratum: [[published erratum appears in Science 1999 Jan
1:283(5398):351]]
2 (bases 1 to 42467)
Watson, R., Bentley, D. and Gattung, S.
The sequence of C. elegans cosmid f26f12
unpublished (1999)
3 (bases 1 to 42467)
Waterson, R.
Direct Submission
Submitted (17-APR-1996) Robert Waterson
4 (bases 1 to 42467)
Waterson, R.
Direct Submission
Submitted (08-JUL-1998) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 42467)
Waterson, R.
Direct Submission
Submitted (02-JUN-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Jul 8, 1998 this sequence version replaced gi.11280109.
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: tw@nematoe.wustl.edu and jesus@sanger.ac.uk

REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
REFERENCE
AUTHORS
JOURNAL
TITLE
JOURNAL
COMMENT

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by

QY 1 RAARTGCCAARTCDATRTGRA 23
 DB 29885 AAAATGCCAAAATGATAGAAA 29863

RESULT 4

LOCUS CELM01G5 37675 bp. DNA INV 21-JUL-1998
 DEFINITION Caenorhabditis elegans cosmid M01G5.
 ACCESSION AF078786
 MID g3329634
 VERSION AF078786.1 GI:3329634
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE 1 (bases 1 to 37675)
 AUTHORS Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laister, N., Latreille, P., Lightning, J., Lloyd, C., McMurry, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifkin, L., Roopra, A., Saunders, D., Showkneen, R., Smalton, N., Smith, A., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstein, L., Wilkinson-Spratt, J. and Mohlman, P.

TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans

JOURNAL Nature 368 (6466), 32-38 (1994)
 MEDLINE 94150718
 REFERENCE 2 (bases 1 to 37675)
 AUTHORS Geisel, C., Kramer, J. and Twyman, B.
 TITLE The sequence of C. elegans cosmid M01G5
 JOURNAL Unpublished (1998)
 REFERENCE 3 (bases 1 to 37675)
 AUTHORS Waterston, R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUL-1998) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1HQ, England
 e-mail: twenematode.wustl.edu and jens@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

The 5' cosmid is Y71D11. Actual start of this cosmid is at base position 1 of CELM01G5; actual end is at 37675 of CELM01G5.

NOTES:

Coding sequences below are predicted from computer analysis, using

the program GeneFinder.P. Green and L. Hillier, ms in preparation).
 Location/Qualifiers

FEATURES

source

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

CDS

gene
 CDS
 join(2941..3663,5495..5838,7824..8064,8594..9213,
 10251..10343,11019..11322,113120..13464)
 /gene="M01G5.1"
 /codon_start=1
 /evidence=not-experimental
 /protein_id="AAC26941.1"
 /db_xref="GI:3329635"

```

/db_xref="PID:93329638"
/db_xref="GI:3329638"
/translation="GMSNRSNDYLAVIGTACGSCFMRPELYPQHGAAFLPYLTC
MCLASLPMFPMENVLGOFSSSAAISVWKVPLRKIGFAOYVTSGFVAFNITSAM
TLFLINSEFSIPCNVDLMDGVDGTLNMTGLGVLACWLAHVECLFQGVKSSGG
KVVAVAVLPPIILTLVLTLLTLDGSLAAVEFLPLKEMILMDLHVGEAAVAFYS
VSSCGGLFTIASKSRFNNTIKDMLIVDVYSLVGLCTFSAGTCEFAISID
DKFHRDGFHLVPEFLAEALAGVAVAPLADFLIMLVYHAOMFVENTISSIDC
EXPLNRNRHRRHVLTVCALITLISIPCLSSGFLPMELLQPIVLPNVPYIARLECM
AINWYGVNDLNDNRKMTVGIMPCOITPMKLFETPCVYVLAIFCEFLMDMSIOYE
SYQPPYMSILWAMCIASPLDIPVIGTMOGCIKAGITQWAMVLYPDDMGPMALH
HRAKRFPLQIPEARLLPPEVEIASRSGVQLVDHMERKLRSYFTPRV"

BASE COUNT      12631 a      6500 c      6589 g      11955 t
ORIGIN

Query Match      79.1%; Score 18.2; DB 37; Length 37675;
Best Local Similarity 65.2%; Pred. No. 96;
Matches 15; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY      1 RAARTGCCRAATCATCTRTGAA 23
      :11:1111:11:1:11:11:11
      Db 24083 AAAGTCCAAATTATATGAA 24105

RESULT 5
CEY47H10/c      CEX47H10      297070 bp      DNA      HTG      29-MAY-1999
LOCUS      Caenorhabditis elegans chromosome I clone Y47H10, WORKING DRAFT
DEFINITION
SEQUENCE, in unordered pieces.
295311
94938508
295311.9      GI:4938508
HTG: HTGS_PHASE1.
Caenorhabditis elegans.
Caenorhabditis elegans.
ORGANISM
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.
1 (bases 1 to 297070)
McLay, K.
Direct Submission
Submitted (29-MAY-1999) Nematode Sequencing Project, Sanger Centre,
Hinxton, Cambridge CB10 1HQ, UK and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jeesanger.ac.uk or rtw@nemastode.wustl.edu
On Jun 1, 1999 this sequence version replaced gi:4914472.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
      source
      Location/Qualifiers
      1..297070
      /organism="Caenorhabditis elegans"
      /db_xref="taxon:6239"
      /chromosome="I"
      /clone="Y47H10"

BASE COUNT      93108 a      55103 c      53632 g      92826 t      2401 others
ORIGIN

Query Match      75.7%; Score 17.4; DB 34; Length 297070;
Best Local Similarity 65.2%; Pred. No. 2,5e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

OY      1 RAARTGCCRAATCATCTRTGAA 23
      :11:1111:11:11:11:11

```

```

DB 287208  GAAATGCCAAATCCCTATGCA 287186

RESULT      6
LOCUS       CEC47F8      21816 bp      DNA      INV      23-NOV-1998
DEFINITION  Caenorhabditis elegans cosmid C47F8, complete sequence.
ACCESSION   AL009246
VERSION     93217213
KEYWORDS    GI:3217213
SOURCE      HTG.
ORGANISM    Caenorhabditis elegans.
            Caenorhabditis elegans.
            Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
            Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
            1 (bases 1 to 21816)
REFERENCE   1
  AUTHORS   McLAY, K.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-DEC-1997) Louis, MO 63110, USA. E-mail:
            joesanger.ac.uk or rvenematode.wustl.edu
  REFERENCE 2 (bases 1 to 21816)
  AUTHORS   Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berk, M.,
            Bonfield, J., Burton, J., Connell, M., Cosey, T., Cooper, J.,
            Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
            Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
            Johnston, L., Jones, M., Kershaw, J., Kirten, J., Laister, N.,
            Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B.,
            O'Callaghan, M., Parsons, J., Percy, C., Riffen, L., Roopra, A.,
            Saunders, D., Showkeen, R., Smalton, N., Smith, A., Sonhammer, E.,
            Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
            Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
            Wilkison-Sproat, J. and Wohlman, P.
  TITLE     2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
            elegans
  JOURNAL   Nature 368 (6466), 32-38 (1994)
  MEDLINE   94150718
  COMMENT   On Jun 13, 1998 this sequence version replaced gi:2664249.
            Coding sequences below are predicted from computer analysis, using
            predictions from Genefinder (P. Green, U. Washington), and other
            available information.
            For a graphical representation of this sequence and its analysis
            see: -
            http://webace.sanger.ac.uk/cgi-
            bin/display7db-wormaceclass-Sequence6object-C47F8
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            IMPORTANT: This sequence is NOT necessarily the entire insert of
            the specified clone. It may be shorter because we only sequence
            overlapping sections once, or longer because we arrange for a small
            overlap between neighbouring submissions.
            IMPORTANT: This sequence is not the entire insert of clone C47F8.
            It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone T09E11 is at 21715 in this sequence. The
            true right end of clone F56H6 is at 105 in this sequence. The start
            of this sequence (1..105) overlaps with the end of sequence 281553
            The end of this sequence (21715..21816) overlaps with the start of
            sequence 281147.
            Location/Qualifiers
              1..21816
              /organism="Caenorhabditis elegans"
              /db_xref="taxon:6239"
              /chromosome="I"
              /clone="C47F8"
              complement(1312..2519)
              /gene="C47F8.4"
              complement(join(1312..1413,1468..1724,2073..2401,
              2461..2519))
              /gene="C47F8.4"
FEATURES
source
gene
CDS

```

/codon_start=1
/protein_id="CA15837.1"
/db_xref="PID:e1344648"
/db_xref="PID:g3875099"
/db_xref="GI:3875099"
/db_xref="SPTREMBL:062113"
/translation="MVIIVPTFKRIIRIPDPTLANTLAHVENLHMLVVEDGYIV
EVRMLEERTNSTIYMAKTRAKGIPRSKQVTRTALIRSSSAKITLQKNGALVYF
ADDDNAIDVRLFTQYIRKVNNTLVGAWGLVGVVAVKVVNQVTAIVRWALSRR
ADWAGFAINLKLINSDAVGTDCRREGAPETCLLMDGLKMDIEIPFGYDAIKVR
DIMVWHTSPTELEIQDQPIDSLGYEVEY"
3341..4845
/gene="C47F8.6"
join(3341..3395,3442..3540,3595..3956,4139..4338,
4383..4509,4569..4658,4705..4845)
/gene="C47F8.6"
/codon_start=1
/protein_id="CA15839.1"
/db_xref="PID:e1344650"
/db_xref="PID:g3875101"
/db_xref="GI:3875101"
/db_xref="SPTREMBL:062115"
/translation="MPRENRLLSVLLPIITFFMFLCLKSEKVEEPLADEAPKONC
TKNGMIGSTAPETODEGSEPMISPAIOKNTVLTLPITETSEKDILMTIVASRD
SYARNIRROTWMKNSSEIYANGRMKSLFVGLARADYKRYKMMQAKIXGDIIV
DMDDTYEELITKSLMIFLFGVSKAPQIKITIKIDEDLMEFPDKMLALTEQGIIDSTV
SLYGLVTPAGRIDFKDITNRYVPESAYSCSQPAYLSGMTYMATREAAOMLSTKH
RDFIOVEDVLLTGILAEEDLGIPRDMPTLYKFPDHKKNDKDIILAHMKNINPYLN
EFMKGLARYSEKGLTRSTSG"
6088..8053
/gene="C47F8.5"
join(6088..6163,6222..6583,7339..7538,7593..7719,
7765..7854,7910..8053)
/gene="C47F8.5"
/codon_start=1
/protein_id="CA15838.1"
/db_xref="PID:e1344649"
/db_xref="PID:g3875100"
/db_xref="GI:3875100"
/db_xref="SPTREMBL:062114"
/translation="MFKILHIMIOJKNELFSKRIARGAOTTAERTMDYGSFIISFA
DIOKTYKLYLPETDIDIVRSPDILMNVASRTDSFARAVLRKTYMNNKNSVILVSDG
KALFLVGVSESDYRRRTVMEAKITGMYVIDEDTDDLPKSLSLILAAVSAPE
FKVIGKIDEDVWFPPDKLIPLDGKVIDPDAAAFYGLLKEGEPIYTKKDAHWYPDY
AYNCTGPAYVAGPFYLATRKAKVILFTKFOFMGTEDSLITGILANDIGIPKKNL
EHVYRHDIQDNQEGKEILAWHSFKNNIPYRNFKVYARHREQNLSMLEINR"
complement(10441..11001)
/gene="C47F8.7"
complement(join(10441..10661,10705..10797,10848..11001))
/gene="C47F8.7"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CA15834.1"
/db_xref="PID:e1344645"
/db_xref="PID:g3875096"
/db_xref="GI:3875096"
/db_xref="SPTREMBL:062110"
/translation="MLPAPATFLILISIAVMHVOCOPGPFVNNKTAADYLCKASKNKY
ILAVNSKDAIYEFETIKSYESETPSDDMMNLPEKVAATSVATSVLPHTSKNI
TRYKQALREGYGEIETAIPEKHSRRVPOLVYICKIRHFSFIPLNDML"
complement(12091..13485)
/gene="C47F8.8"
complement(join(12091..12144,12193..12267,12332..12533,
12584..12810,12859..13034,13207..13272,13317..13485))
/gene="C47F8.8"
/note="similar to zinc finger, C4 type (two domains)"
/codon_start=1
/protein_id="CA15840.1"
/db_xref="PID:e1344651"
/db_xref="PID:g3875102"
/db_xref="GI:3875102"
/db_xref="SPTREMBL:062116"
/translation="MGLISKNGPCOVCHNTETTRHFGIISCTACAAFFRSISGMQY

ICTAENRCTISFDRKFRACRYASCIAKAKLEHSMSPATSEGOSSSKTDERSD
PEYSKILKMHDDLLQYVKEKVMKMKRRCERNTQITITISDMLITTRQSPKM
EELCRPGVDLEKEDEVALRYRFSNIWDSVLISVLSAGDSSTDDDRSLR
IOEVKSTGSSIAHLNINLYEFESFAFKALCIKWLKLEFSPKQKIIAHSHBGVIGALR
NYENCTELADMEIAIRIGDTITQIVTVQNTYHHEVRLYQIGMPL"
15076..17404
/gene="C47F8.2"
join(15076..15091,15138..15365,16255..16831,17007..17208,
17252..17404)
/gene="C47F8.2"
/note="similar to zinc finger, C4 type (two domains)"
/codon_start=1
/protein_id="CA15836.1"
/db_xref="PID:e1344647"
/db_xref="PID:g3875098"
/db_xref="GI:3875098"
/db_xref="SPTREMBL:062112"
/translation="MSRSEPNRSKICIGSICTAKFPHRGLSCNACAFERRVSLNT
RYLKNNSNCTISPGQRTFCACRSCVDAKMRDVOORSTNATPKITLNTSG
SSNREVVGLSTISRTISTFTGFTDSNSHLVQDHLNIONNENCKMKRYSDL
LEISHMNLRRYVQVPEPEKELVSTLSNLAVLFPQNIILQIKSINDLMTCAVNG
SAMESCVCPGTGKLSQEDVRLQKSFQFATWDSITQWYSIATDIDSEVSGESL
SOFISHISTIGACILANKLIDYEPFAARSCVWSIGAGATESKRYVAORHEAVAS
ALKMYKTESMSDSEIATRIEELSFGRSSDVQMIYDITIKLYKIGIKPLFPGN"
19133..20530
/gene="C47F8.1"
join(19133..19209,19717..20340,20386..20530)
/gene="C47F8.1"
/note="predicted using GeneFinder: cDNA EST YK22967.5
comes from this gene; cDNA EST YK22967.3 comes from this
gene"
/codon_start=1
/protein_id="CA15835.1"
/db_xref="PID:e1344646"
/db_xref="PID:g3875097"
/db_xref="GI:3875097"
/db_xref="SPTREMBL:062111"
/translation="MSSFQLOQLPSLPIDKEYFKSLGVCDLILNCSSNKSRLVYKSL
KDITFTLIVAEKRYKIMLKEDPSKHKARARANSNGTSLEKNIKRMHRTVEYV
LDCNTVEGDVYKVEYLLKKEPYLKFECPMISDLMQIMNCADVILIISYDR
VKVFDNVPILFNECITLHGKFSKGMNLFTWOSKREKRTLOCVALTERLIGPMK
KQANEKVLKIDVESYKRIIFPKKDKIKKSTTYNLYKNDGELATLAKFKQELCFH
TENDR"
complement(613..679),complement(487..564),
complement(70..431),complement(281553.1.36172..36371),
complement(281553.1.36000..36126),
complement(281553.1.35864..35953),
complement(281553.1.35704..35817),
complement(281553.1.35704..35817)
/gene="C47F8.3"
join(complement(613..679),complement(487..564),
complement(70..431),complement(281553.1.36172..36371),
complement(281553.1.36000..36126),
complement(281553.1.35864..35953),
complement(281553.1.35704..35817))
/gene="C47F8.3"
/codon_start=1
/protein_id="CA15841.1"
/db_xref="PID:e1344652"
/db_xref="PID:g3875103"
/db_xref="SPTREMBL:062117"

Query Match 75.7%: Score 17.4; DB 36; Length 21816;
Best Local Similarity 65.2%: Pred. No. 2.2e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;
QY 1 RAATGCCCAARTCDATRTGAA 23
Db 1174 GAATGCCCAAAATCCCTATGGAA 1152

RESULT 7
LOCUS HS992D9
DEFINITION Human DNA sequence from BAC 992D9 on chromosome 22q12.1 contains


```

variation      /note="LM3 repeat: matches 909. .150 of consensus"
                9630. .9632
                /note="clone 90L6; GTC in this entry; substitution"
                /replace="ggc"
                10126. .10128
                /note="clone 90L6; CAG in this entry; substitution"
                /replace="cgg"
                10293. .10361
                /note="MIR repeat: matches 77. .144 of consensus"
                10337. .10339
                /note="clone 90L6; TCA in this entry; substitution"
                /replace="tta"
                11450. .11545
                /note="MIR repeat: matches 49. .145 of consensus"
                11956. .12000
                /note="MIR2 repeat: matches 146. .102 of consensus"
                12714. .12790
                /note="MIR2 repeat: matches 47. .124 of consensus"
                12903. .12936
                /note="U2 repeat: matches 34. .1 of consensus"
                13166. .13546
                /note="LM35 repeat: matches 535. .921 of consensus"
                13242. .13244
                /note="clone 90L6; ACA in this entry; substitution"
                /replace="ata"
                13656. .13658
                /note="clone 90L6; TGT in this entry; substitution"
                /replace="ttt"
                13916. .13918
                /note="clone 90L6; CAA in this entry; substitution"
                /replace="cca"
                14086. .14088
                /note="clone 90L6; AGA in this entry; substitution"
                /replace="aaa"
                14151. .14153
                /note="clone 90L6; GTG in this entry; substitution"
                /replace="ggg"
                14284. .14286
                /note="clone 90L6; GTG in this entry; substitution"
                /replace="ggc"
                14609. .14896
                /note="AluY repeat: matches 295. .7 of consensus"
                15373. .15666
                /note="AluX repeat: matches 293. .1 of consensus"
                15615. .15617
                /note="clone 90L6; CCT in this entry; substitution"
                /replace="ctt"
                15798. .15800
                /note="clone 90L6; AGG in this entry; substitution"
                /replace="aag"
                15814. .16224
                /note="MIR1B repeat: matches 390. .1 of consensus"

variation      74.88; Score 17.2; DB 9; Length 155147;
Best Local Similarity 70.08; Pred. No. 3e+02;
Matches 14; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY      4 RTGCCAARTCDATRTGAA 23
        :||||:||||:||||:||||:
Db 100531 ATGCCAAGTCAATGTGCA 100550

RESULT      8
AF067844    218336 bp    DNA    PRI    09-FEB-1999
LOCUS      *Homo sapiens chromosome 10 clone PTEN, complete sequence.
ACCESSION  AF067844
VERSION    94240386
KEYWORDS   AF067844.1 GI:4240386
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

```

REFERENCE
AUTHORS      Eutheria; Primates; Catarrhini; Homiinae; Homo.
              1 (bases 1 to 218336)
              Jensen,K., de la Bastide,M., Parsons,R., Parnell,L.D., Dedhia,N.,
              Gottesman,T., Gnoj,L., Kaplan,N., Lodhi,M., Johnson,A.F.,
              Shohdy,N., Hasegawa,A., Haberman,K., Huang,E.N., Schutz,K.,
              Calma,C., Granat,S., Wiegler,M. and McCombie,W.R.
              Genomic sequence of PTEN/MMAC1
              Unpublished
              2 (bases 1 to 218336)
              Jensen,K., de la Bastide,M., Parsons,R., Parnell,L.D., Dedhia,N.,
              Gottesman,T., Gnoj,L., Kaplan,N., Lodhi,M., Johnson,A.F.,
              Shohdy,N., Hasegawa,A., Haberman,K., Huang,E.N., Schutz,K.,
              Calma,C., Granat,S., Wiegler,M. and McCombie,W.R.
              Direct Submission
              Submitted (18-MAY-1998) Lita Annenberg Hazen Genome Sequencing
              Center, Cold Spring Harbor Laboratory, 1 Bungtown Rd., Cold Spring
              Harbor, NY 11724, USA

TITLE
JOURNAL
REFERENCE
AUTHORS

FEATURES
SOURCE
1. .218336
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="10"
   /clone="PTEN"
1. .106991
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="BAC 265N13"
   /chromosome="10"
22308. .22338
   /gene="PTEN"
   /note="5'-UTR defined by comparison to PTEN cDNA U93051"
   join(22308. .23417,51995. .52079,83482. .83526,89015. .89058,
   90987. .91225,110086. .110227,115821. .115987,118862. .119086,
   123258. .124345)
   /gene="PTEN"
   /note="mRNA coordinates delineated by comparison to PTEN
   cDNA U93051"
22308. .124345
   /gene="PTEN"
   /note="the coding region of PTEN, as defined by the cDNA,
   identifies 9 exons within this region; identical to MMAC1
   (U92346) and PTEN (U93051)"
   /evidence="experimental"
22308. .23417
   /gene="PTEN"
   /function="5'-UTR and initial segment of the CDS"
   /number=1
   join(22339. .23417,51995. .52079,83482. .83526,89015. .89058,
   90987. .91225,110086. .110227,115821. .115987,118862. .119086,
   123258. .124343)
   /gene="PTEN"
   /note="coding regions delineated by comparison to PTEN
   cDNA"
   /codon_start=1
   /product="PTEN"
   /protein_id="AA013528.1"
   /db_xref="PID:94240387"
   /db_xref="GI:4240387"
   /translation="MTRAIKEIYSRNRKRRQEGFDLDTIYPTNIAMGFAPRLRG
   VRRNIDVVRFLDSRKHNYKIYNLCAEHYDTAFKNCVAVQYPRDHPDLELIK
   PFCEDLDQWLSEDDNHYVAIHCRAAGRTVMICAYILHGRKFLKQADALFYGEVVT
   RDKRGVITPSORRYVYVSYILKNHLDYRVALLFHMKMFEIIPMFGSGTGNPQFVVC
   OLKVKIYSSNSGPTRRDKMFEFPEFQPLPVGGLDKVEEFHKKRMKLRKDKMFEFVW
   TFLIPGERTSEKVENSLCDDIDISCITERADNCKEYVILLTKNDLDAKADKAN
   RYSPNRYKLYITKYVEPSNPEASSISVTYDVSNDPNDHRYSDTSDSPENPEF
   DEQHTQITKV"
   /gene="PTEN"
   /number=2
   58169. .218336
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /clone="BAC 60C5"

```

FEATURES

Source

Location/Qualifiers

1..97015

/organism="Bacillus subtilis"

/strain="168"

/db_xref="taxon:1423"

/map="325-333"

complement(1..188)

/gene="ipa-0r"

complement(1..175)

/partial

/gene="ipa-0r"

/codon_start=1

/transl_table=1

/protein_id="CAA51556.1"

/db_xref="PID:g413924"

/db_xref="GI:413924"

/db_xref="SWISS-PROT:P39576"

/translation="MRQIRVELTSTKPKPPDPNQLSFGFVTFDHFVNDYADKGM
YDPIRIYQPLSMQ"

complement(181..188)

/gene="ipa-0r"

complement(316..1091)

/gene="ipa-1r"

complement(316..1074)

/gene="ipa-1r"

/codon_start=1

/transl_table=1

/protein_id="CAA51557.1"

/db_xref="PID:g413925"

/db_xref="GI:413925"

/db_xref="SWISS-PROT:P39577"

/translation="MKMTNTVILITGSGAGIGELAKRIELGNEVILIGRSEARLAE
AKQDLPHTKQCDVADRSOREALYEWALKEYPNLVYNNAGIQKIDPKKGEELF
VDGGEILINFOAPVHLSALFPLHMKQPEAIVQVTSGLAFNPPLAVYVPCATKAALH
SFSLTLEHHVBDTSVEYIEEMAPWPDGLNOKSRDQGLTYRGISSEYVQFLDGLK
EGKQETINERVEGRDARTRDYDRLEFQMTQEN"

complement(1084..1091)

/gene="ipa-1r"

complement(1165..2354)

/gene="ipa-2r"

complement(1165..2343)

/gene="ipa-2r"

/codon_start=1

/transl_table=1

/protein_id="CAA51558.1"

/db_xref="PID:g413926"

/db_xref="GI:413926"

/db_xref="SWISS-PROT:P39578"

/translation="MKRRFGPIITAIILFAGALAIPTSSWLTFGITDKRKRESATLALN
PMSHGILLDDQMLKDEPTVLPITGSSSLFLDEHPSPNRYOVNNEGTPPLVKGSSQ
SLIHSLFNAHMDQLKQRKIVFLVSPQWFKRGSDEGHPAPNSALOGDLAFQDQID
PEIKKAKMRMLRFAVONDAILSELYKAWNGQOTVWNLKPAKVVYLMLEKDDY
YSTTESGPKRYISQSYOKDKSMSELNLKADQSGRHSNSDFHIDNPVYKKLPKPYK
LKGNKGRSYAASPEYGDPEMLDILKDAGAEPMFTIIPVNGKMYDVTGPKKGRDYY
YKYNKQIRAKGFOVADFSGHEVDPEFMKDTIHIGKMGVYVDKALIDELFKTKGTYS

gene

CDS

complement(2343..2596)

/gene="ipa-3r"

complement(2343..2579)

/gene="ipa-3r"

/codon_start=1

/transl_table=1

/protein_id="CAA51559.1"

/db_xref="PID:g413927"

/db_xref="GI:413927"

/db_xref="SWISS-PROT:P39579"

/translation="MDFKQEVLDVLAECODDIVKENPDIETFEEDGLDSFGVEILL
AINRDIIVPIETFPDWNTPNNVYNQSELK"

complement(2349..2354)

/gene="ipa-2r"

complement(2589..2966)

/gene="ipa-3r"

gene

CDS

complement(1165..2343)

/gene="ipa-2r"

complement(1165..2343)

/gene="ipa-2r"

/codon_start=1

/transl_table=1

/protein_id="CAA51558.1"

/db_xref="PID:g413926"

/db_xref="GI:413926"

/db_xref="SWISS-PROT:P39578"

/translation="MKRRFGPIITAIILFAGALAIPTSSWLTFGITDKRKRESATLALN
PMSHGILLDDQMLKDEPTVLPITGSSSLFLDEHPSPNRYOVNNEGTPPLVKGSSQ
SLIHSLFNAHMDQLKQRKIVFLVSPQWFKRGSDEGHPAPNSALOGDLAFQDQID
PEIKKAKMRMLRFAVONDAILSELYKAWNGQOTVWNLKPAKVVYLMLEKDDY
YSTTESGPKRYISQSYOKDKSMSELNLKADQSGRHSNSDFHIDNPVYKKLPKPYK
LKGNKGRSYAASPEYGDPEMLDILKDAGAEPMFTIIPVNGKMYDVTGPKKGRDYY
YKYNKQIRAKGFOVADFSGHEVDPEFMKDTIHIGKMGVYVDKALIDELFKTKGTYS

gene

CDS

complement(2343..2596)

/gene="ipa-3r"

complement(2343..2579)

/gene="ipa-3r"

/codon_start=1

/transl_table=1

/protein_id="CAA51559.1"

/db_xref="PID:g413927"

/db_xref="GI:413927"

/db_xref="SWISS-PROT:P39579"

/translation="MDFKQEVLDVLAECODDIVKENPDIETFEEDGLDSFGVEILL
AINRDIIVPIETFPDWNTPNNVYNQSELK"

complement(2349..2354)

/gene="ipa-2r"

complement(2589..2966)

/gene="ipa-3r"

gene

CDS

complement(1165..2343)

/gene="ipa-2r"

complement(1165..2343)

/gene="ipa-2r"

/codon_start=1

/transl_table=1

/protein_id="CAA51558.1"

/db_xref="PID:g413926"

/db_xref="GI:413926"

/db_xref="SWISS-PROT:P39578"

/translation="MKRRFGPIITAIILFAGALAIPTSSWLTFGITDKRKRESATLALN
PMSHGILLDDQMLKDEPTVLPITGSSSLFLDEHPSPNRYOVNNEGTPPLVKGSSQ
SLIHSLFNAHMDQLKQRKIVFLVSPQWFKRGSDEGHPAPNSALOGDLAFQDQID
PEIKKAKMRMLRFAVONDAILSELYKAWNGQOTVWNLKPAKVVYLMLEKDDY
YSTTESGPKRYISQSYOKDKSMSELNLKADQSGRHSNSDFHIDNPVYKKLPKPYK
LKGNKGRSYAASPEYGDPEMLDILKDAGAEPMFTIIPVNGKMYDVTGPKKGRDYY
YKYNKQIRAKGFOVADFSGHEVDPEFMKDTIHIGKMGVYVDKALIDELFKTKGTYS

gene

CDS

complement(2343..2596)

/gene="ipa-3r"

complement(2343..2579)

/gene="ipa-3r"

/codon_start=1

/transl_table=1

/protein_id="CAA51559.1"

/db_xref="PID:g413927"

/db_xref="GI:413927"

/db_xref="SWISS-PROT:P39579"

/translation="MDFKQEVLDVLAECODDIVKENPDIETFEEDGLDSFGVEILL
AINRDIIVPIETFPDWNTPNNVYNQSELK"

complement(2349..2354)

/gene="ipa-2r"

complement(2589..2966)

/gene="ipa-3r"

gene

CDS

complement(1165..2343)

/gene="ipa-2r"

complement(1165..2343)

/gene="ipa-2r"

/codon_start=1

/transl_table=1

/protein_id="CAA51558.1"

/db_xref="PID:g413926"

/db_xref="GI:413926"

/db_xref="SWISS-PROT:P39578"

/translation="MKRRFGPIITAIILFAGALAIPTSSWLTFGITDKRKRESATLALN
PMSHGILLDDQMLKDEPTVLPITGSSSLFLDEHPSPNRYOVNNEGTPPLVKGSSQ
SLIHSLFNAHMDQLKQRKIVFLVSPQWFKRGSDEGHPAPNSALOGDLAFQDQID
PEIKKAKMRMLRFAVONDAILSELYKAWNGQOTVWNLKPAKVVYLMLEKDDY
YSTTESGPKRYISQSYOKDKSMSELNLKADQSGRHSNSDFHIDNPVYKKLPKPYK
LKGNKGRSYAASPEYGDPEMLDILKDAGAEPMFTIIPVNGKMYDVTGPKKGRDYY
YKYNKQIRAKGFOVADFSGHEVDPEFMKDTIHIGKMGVYVDKALIDELFKTKGTYS

gene

CDS

complement(2343..2596)

/gene="ipa-3r"

complement(2343..2579)

/gene="ipa-3r"

/codon_start=1

/transl_table=1

/protein_id="CAA51559.1"

/db_xref="PID:g413927"

/db_xref="GI:413927"

/db_xref="SWISS-PROT:P39579"

/translation="MDFKQEVLDVLAECODDIVKENPDIETFEEDGLDSFGVEILL
AINRDIIVPIETFPDWNTPNNVYNQSELK"

complement(2349..2354)

/gene="ipa-2r"

complement(2589..2966)

/gene="ipa-3r"

gene

CDS

complement(1165..2343)

/gene="ipa-2r"

complement(1165..2343)

/gene="ipa-2r"

/codon_start=1

/transl_table=1

/protein_id="CAA51558.1"

/db_xref="PID:g413926"

/db_xref="GI:413926"

/db_xref="SWISS-PROT:P39578"

/translation="MKRRFGPIITAIILFAGALAIPTSSWLTFGITDKRKRESATLALN
PMSHGILLDDQMLKDEPTVLPITGSSSLFLDEHPSPNRYOVNNEGTPPLVKGSSQ
SLIHSLFNAHMDQLKQRKIVFLVSPQWFKRGSDEGHPAPNSALOGDLAFQDQID
PEIKKAKMRMLRFAVONDAILSELYKAWNGQOTVWNLKPAKVVYLMLEKDDY
YSTTESGPKRYISQSYOKDKSMSELNLKADQSGRHSNSDFHIDNPVYKKLPKPYK
LKGNKGRSYAASPEYGDPEMLDILKDAGAEPMFTIIPVNGKMYDVTGPKKGRDYY
YKYNKQIRAKGFOVADFSGHEVDPEFMKDTIHIGKMGVYVDKALIDELFKTKGTYS

gene

CDS

complement(2343..2596)

/gene="ipa-3r"

complement(2343..2579)

/gene="ipa-3r"

/codon_start=1

/transl_table=1

/protein_id="CAA51559.1"

/db_xref="PID:g413927"

/db_xref="GI:413927"

/db_xref="SWISS-PROT:P39579"

/translation="MDFKQEVLDVLAECODDIVKENPDIETFEEDGLDSFGVEILL
AINRDIIVPIETFPDWNTPNNVYNQSELK"

complement(2349..2354)

/gene="ipa-2r"

complement(2589..2966)

/gene="ipa-3r"

gene

CDS

complement(1165..2343)

/gene="ipa-2r"

complement(1165..2343)

/gene="ipa-2r"

/codon_start=1

/transl_table=1

/protein_id="CAA51558.1"

/db_xref="PID:g413926"

/db_xref="GI:413926"

/db_xref="SWISS-PROT:P39578"

/translation="MKRRFGPIITAIILFAGALAIPTSSWLTFGITDKRKRESATLALN
PMSHGILLDDQMLKDEPTVLPIT

QY	1	RAARTGCCRAARTCDATRTGRPA	23
Db	21697	GAATGCCGAATGCTTTGAAA	21675
Query Match			73.9%
Best Local Similarity			65.2%
Matches	15	Conservative	6
		Mismatches	2
		Indels	0
		Gaps	0
RESULT	10		
BSDB0020			
LOCUS			
DEFINITION			
ACCESSION			
NID			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			

[illegible]

Viari, A., Mambitt, R., Wedler, E., Wedler, H., Weltzenegger, T.,
Winters, P., Wipacit, A., Yamamoto, H., Yamae, K., Yasumoto, K., Yata, K.,
Yoshida, K., Yoshikawa, H. F., Zumbstein, E., Yoshikawa, H. and
Danchin, A.

TITLE	JOURNAL	YEAR	PAGE
The complete genome sequence of the gram-positive bacterium <i>Bacillus subtilis</i>	Nature	390 (6657)	249-256 (1997)
00044033			

JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL

Nature 390 (6657), 249-256 (1997)
98044033
2 (bases 1 to 212150)
Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.
Direct Submission
Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur,
Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75722
Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr,
adanchin@pasteur.fr. Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45
68 89 48

```

source      1. 212150
            Location/Qualifiers
            /organism="Bacillus subtilis"
            /strain="168"
            /db_xref="taxon:1423"
            complement(7. .991)
            /gene="ywkB"
            complement(7. .23)
            /gene="ywkB"
            complement(32. .991)
            /gene="ywkB"
            /function="unknown"
            /note="similar to hypothetical proteins"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB15732.1"
            /db_xref="pid:e1186203"
            /db_xref="pid:g2636241"
            /db_xref="GI:2636241"
            /db_xref="SWISS-PROT:P45869"
            /translation="MSIDILIPFFVVLGWFAGHFGSYDAKSAGKSYLTVTYK
            ALPAPIFGILTTSSEFLSOVPLMISLIGIVGPIIILGCHPIRKYDSTNSVS
            LNSAPDAPFNGCIPYGLSGEGNENVAIPATGTGVYNAIDPLAIITATGESSKNE
            ESGDSFMRKGTGSLIHGCEPAPAPLISMILVFNFTLPDELGRKMDQGSTTSV
            ALFAGVAVGIRKIKLSMPALGIALIKVAQVPAIMPALPLALAIGLPADQTKAILLVAF
            PESAAMAIARFAEEQDEETAVAFVSAIILSLIPITIALTA"
            complement(1073. .2821)
            /gene="ywkA"
            complement(1073. .2821)
            /gene="ywkA"
            /function="unknown"
            /note="similar to malate dehydrogenase"
            /codon_start=1
            /transl_table=11
            /protein_id="CAB15733.1"
            /db_xref="pid:e1186206"
            /db_xref="pid:g2636242"
            /db_xref="GI:2636242"
            /db_xref="SWISS-PROT:P45868"
            /translation="MGYLLTWLITISRKKEICLNINIKRKEGHELTTLGRKEVISIPTL
            NKGVFSEERRELEGLLEPTVLSLQOQARAEQOAPDRLRQNYVLSDLAANN
            EYLFEKLKNHIREMLPVYAPPTGEALAEQESHYRRPGQYISIDNIDIGEKAFENL
            HATADIDILVYATDSESIIGIDMGVSGEINATIGRLAVTTAAGIDPSRYIPVYLDL
            TNNHELNDPLITGNKRRVOGERTAEIDAVYKRAALKEFKALLHEDLGNNKNAKI
            MKRYKHEILFENDDIQGTGATTLAGVLAAMKGTASIKDQVNVIFGASGIGIADQI
            RDTVAIAGISEEANKRFTYLDIGRLITLEDIEGLDFOKPYLRNADSVKMKRDEKQO
            IPEDVAVOAKPTLIGTSGVGAFTETIEKMAHSVDRPIVIMNSNPTHLAEAVPDD
            LEKTKDGVLAITATGSPFNVEYNGVSYETIGOSNNAFAPPGIGSGISVAERITTPAF
            AATNAALAEMLDTPRPGAGLPSIDKLOEVSIOVALVAEALITDGVANRQPEDVYQCA
            VDAAMVTEIRKVIAMK"
            complement(3011. .3647)
            /gene="cdk"
            complement(3011. .3031)
            /gene="cdk"
            complement(3060. .3647)
            /gene="cdk"
            /EC_number="2.7.1.21"

```

```

/codon_start=1
/transl_table=11
/product="thymidine kinase"
/protein_id="CAB15734.1"
/db_xref="PID:e1186207"
/db_xref="PID:g2636243"

```

[illegible]

```

complement(5714. .5735)
/gene="yw1i"
complement(5714. .6710)
/gene="yw1i"
complement(5745. .6710)
/gene="yw1i"
/function="unknown"
/note="similar to glycerol-inducible protein"
/transl_start=1
/codon_start=1
/protein_id="CAB15737.1"
/db_xref="PID:el186520"
/db_xref="PID:g2636246"
/db_xref="GI:2636246"
/db_xref="SWISS-PROT:Q03224"
/translacion="MESISLMEVRYTAEFAALASARMMRGPKKDEADEAATSAARVDYV
DTPVAGGVVIGDEEDAEALNLTIGKLLNGGSPVDVAVPLESTNLISGAGNALT
VIAADHSTILNMPDYMOKIANGPAVACSIDIEPVIIDNLKAYAKAKKHVEYVAT
VILNHRKAKITIGLPPGESEAIKILINGDVAGIINTAFHGTVDILFGSGAPESVLSA
VALALAGEIIGLRLKREGEIIRKRMELIDSKVLRNEDLVKGDGDAIFATGTDDDE
LLKGVFGSGVGTSTESVVIARAKGTVRFVDGRHSLKKRPNIIRP"
complement(6741. .8030)
/gene="mur2"
complement(6741. .8030)
/gene="mur2"

```

/translation="MYTMRSGGLEILQSGMSGSGSEELIRVRKATKATQKQVQVAFVPS
 VIDNRYSAAAYSHNGTSGTSTYATSSADAIWPHISSDTPVADVQGFDDQVEYVLS
 SLADGKIVYIAGIDMGIDKREPEGGVYVNPIMAIARESTYKIDQAVCSGSPASRTORLID
 GRPASTYDPTVLIVGAASAYARCRHHHEVPGRSK"
 complement(3698. .3936)
 /gene="rpmB"
 complement(3698. .3720)
 /gene="rpmB"
 complement(3736. .3936)
 /gene="rpmB"
 /codon_start=1
 /translation_table=11
 /product="ribosomal protein L31"
 /protein_id="CAB15735.1"
 /db_xref="PID:e1186208"
 /db_xref="PID:g2636244"
 /db_xref="GI:2636244"
 /db_xref="SWISS-PROT:003223"
 /translation="MKRGCTHDPNKKATVCAQGNPEFGSVAEVRVEIGCECHPFTT
 GRQFASADGQVDEPFNKRYGLK"
 complement(4055. .5338)
 /gene="rho"
 complement(4055. .5338)
 /gene="rho"
 /codon_start=1
 /translation_table=11
 /product="transcriptional terminator Rho"
 /protein_id="CAB15736.1"
 /db_xref="PID:e1186209"
 /db_xref="PID:g2636245"
 /db_xref="GI:2636245"
 /db_xref="SWISS-PROT:003222"
 /translation="MKRVYSSESNNKLELYELARHKISYISKLTKEELPAILAKA
 NAEQDLDLMGVELEIIQSGSEFLRPINYSFSSSDIYISAQIRFPLRNGDVSKR
 VRRPEEMREYRGGLIHYAVNGDDPESAKERVHFPALTPLPDROMVLETKKNEFLSTRI
 MDMAAPVQFGRCGLIHYAPAGATMLKEIANSITANOPEALIVLIDPEPEVTDID
 ERSVAGVSTGEPEPEENIKVAEVLERRARLVHKKYVILMDSITLARAANYIN
 IIPSGRILSGIDIPAAVHRPRRFGGAANIIEGSGITLATAVDGSRMDYIYEFEF
 KGTGMELHEDRSIAEKRIPTAIDIRSGTRKEELVPREHLRLMSIRKTMSSDPDT
 AEKFFRRKKTKTKNQEFILINEMKQANLSARR"
 complement(5714. .5735)
 /gene="ywj1"
 complement(5714. .6710)
 /gene="ywj1"
 complement(5745. .6710)
 /gene="ywj1"
 /function="unknown"
 /note="similar to glycerol-inducible protein"
 /codon_start=1
 /translation_table=11
 /protein_id="CAB15737.1"
 /db_xref="PID:e1186210"
 /db_xref="PID:g2636246"
 /db_xref="GI:2636246"
 /db_xref="SWISS-PROT:003224"
 /translation="MEKTSISMEYRTAEALASARWGRGKQKDEADPAATSNARQVPE
 DTVMKGVYVIGSEBDEAPMLYIGKRLTNGCGRPVYDVAQVDPLESTNITLISGNNALT
 VTAADGHTLNAPADMTMOKIIVAGPAVGCIDIEAPVIDNTKAVAKANKVDEYVAT
 ILNERRAKITISELRSEGAIRKINDGVAGINAPFHTGVDILFGSGAPRECVLSA
 VALALAGGIIIGLIPQSEEAIRCHRMGLDSTYLRMEDLVKGDPAIFATGFTDDEE
 LLKGQFGSGVGTESIIVAKSGTVRFVDFGHSLLKKPNLYIR"
 complement(6741. .8030)
 /gene="murZ"
 complement(6741. .8030)
 /gene="murZ"

/EC-number="2.5.1.7"
/function="peptidoglycan biosynthesis"
/note="alternate gene name: murZ, lssF, rev-4"
/codon_start=1
/transl_table=11
/product="UDP-N-acetylglucosamine
1-carboxylinyltransferase"
/protein_id="CA815738.1"
/db_xref="PID:e1186211"
/db_xref="PID:g2636247"
/db_xref="GI:2636247"
/translation="MERT10AGDPSLNGTVHISGAKSAVALPATIANSSEYITIEL
PESIDETLIDLKEIGNHFENGEMVVDPTSMISPLNGKYKRTASAYTGMAAL
GRKQAVTIGPCGCHLPRPIDOHIKFEALCAVTTGQATITRARRLGARTYIDY
VSVAGTINMLAAVLAAGKTTIENAAKEPITIDVATLTSMGAKIKAGAGTVRIDV
KELHGCRTIIPRIENAGETIAGAAAGKEVITIDVITPLESTLALREMGYHTERS

Query Match 73.9%; Score 17; DB 1; Length 212150;
Best Local Similarity 65.2%; Pred. No. 3.8e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGAA 23
:||:||||:||||:||||:||||:
Db 136262 GAATGCCGAATCGTTTGAA 136284

RESULT 11
YSCMET10A/c
LOCUS YSCMET10A 3650 bp DNA PLN 08-FEB-1995
DEFINITION Saccharomyces carlsbergensis assimilatory sulfite reductase (MET10)
ACCESSION L26504
NID 9498310
VERSION L26504.1 GI:498310
KEYWORDS MET10 gene.
SOURCE Saccharomyces carlsbergensis DNA.
ORGANISM Saccharomyces pastorianus
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
REFERENCE 1 (bases 1 to 3650)
Hansen, J., Charest, H. and Kiehlend-Brandt, M.C.
Two divergent MET10 genes, one from Saccharomyces cerevisiae and
one from Saccharomyces carlsbergensis, encode the alpha subunit of
sulfite reductase and specify potential binding sites for FAD and
NADPH
NADPH
J. Bacteriol. 176 (19), 6050-6058 (1994)

COMMENT
MEDLINE 95014040
REFERENCE 2 (bases 1 to 3650)
AUTHORS Hansen, J.
TITLE Direct Submission
Submitted (03-DEC-1993) J. Hansen, Carlsberg Research Laboratory,
Gamle Carlsbergvej 10, Copenhagen Valby, Denmark, 2500
On Jun 10, 1994 this sequence version replaced gi:432489.
FEATURES
source location/Qualifiers
1. 3650
/organism="Saccharomyces pastorianus"
/db_xref="taxon:27292"
215..3322
/gene="MET10"
215..3322
/EC-number="1.8.1.2"
/note="flavin-binding subunit"
/codon_start=1
/product="assimilatory sulfite reductase"
/protein_id="AA61981.1"
/db_xref="PID:g432480"
/db_xref="GI:432490"
/translation="MSVEFATNPFGTKATLPSYSTPYATISSVLFENNVDIFARK
SFSQADLHQLKMSLQNSGSKATFEQLDRSAGLTLPLGFSGLNNTATVAP
GFLAPYINSIQTPHAKGFLINAGYNTSGSVINDYATADAAKLNKTVATP
STYENVALALAVATFSNRSGINFDGANYSKTVPLVETAAESSIIIGKSKVIA
PNAFEDVLKFNLSGLRLHNPQYPAQDAETVYVYGSLESELSISGNNAKIG

BASE COUNT 1101 a 743 c 696 g 1110 t
ORIGIN

Query Match 73.9%; Score 17; DB 7; Length 3650;
Best Local Similarity 65.2%; Pred. No. 3.2e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGAA 23
:||:||||:||||:||||:||||:
Db 2247 GAATGCAATTCATGCGAA 2225

RESULT 12
HS19408
LOCUS HS19408 126839 bp DNA PRI 23-MAR-1999
DEFINITION Human DNA sequence from clone 19408 on chromosome 6q24.1-25.3
ACCESSION AL031769
NID 94153934
VERSION AL031769.1 GI:4153934
KEYWORDS HMG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 126839)
Smith, S.
TITLE Direct Submission
Submitted (18-MAR-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Jan 13, 1999 this sequence version replaced gi:4007130.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 19408. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome 6, constructed by the Sanger Centre Chromosome 6
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
19408 is from the library Rpg11 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR:pcrYAC2.
FEATURES
source location/Qualifiers
1. 126839
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="19408"

repeat_region	/map:"q24.15.3"	repeat_region	/note="AluY repeat: matches 1. .144 of consensus"
repeat_region	/clone:"lib="RC11"	repeat_region	23681. .23985
repeat_region	/note="AluUb repeat: matches 194. .293 of consensus"	repeat_region	/note="AluSc repeat: matches 1. .297 of consensus"
repeat_region	224. .431	repeat_region	24112. .24313
repeat_region	/note="L2 repeat: matches 2279. .2496 of consensus"	repeat_region	/note="MIR repeat: matches 41. .251 of consensus"
repeat_region	2543. .3170	repeat_region	24716. .25220
repeat_region	/note="L2 repeat: matches 418. .1067 of consensus"	repeat_region	/note="MIR repeat: matches 57. .626 of consensus"
repeat_region	3261. .3461	repeat_region	25365. .25660
repeat_region	/note="L2 repeat: matches 2352. .2539 of consensus"	repeat_region	/note="AluUb repeat: matches 1. .293 of consensus"
repeat_region	5020. .5314	repeat_region	26465. .26504
repeat_region	/note="AluY repeat: matches 1. .295 of consensus"	repeat_region	/note="20 copies 2 mer ct 80% conserved"
repeat_region	5894. .6199	repeat_region	26924. .26969
repeat_region	/note="L1MA7 repeat: matches 5982. .6289 of consensus"	repeat_region	/note="23 copies 2 mer ag 87% conserved"
repeat_region	6200. .6733	repeat_region	27427. .27603
repeat_region	/note="L1PA15 repeat: matches 5595. .6144 of consensus"	repeat_region	/note="AluY repeat: matches 128. .304 of consensus"
repeat_region	6952. .7260	repeat_region	31176. .31483
repeat_region	/note="AluSc repeat: matches 6. .303 of consensus"	repeat_region	/note="AluSc repeat: matches 1. .307 of consensus"
repeat_region	7439. .8958	repeat_region	31485. .31638
repeat_region	/note="L1M4 repeat: matches 2116. .3733 of consensus"	repeat_region	/note="77 copies 2 mer aa 68% conserved"
repeat_region	9186. .10885	repeat_region	31485. .31629
repeat_region	/note="L1M5 repeat: matches 4442. .6173 of consensus"	repeat_region	/note="5 copies 29 mer 71% conserved"
repeat_region	11407. .11518	repeat_region	31905. .32424
repeat_region	/note="L1M1D repeat: matches 1. .107 of consensus"	repeat_region	/note="match: GSS AQJ12911 clone R-1A24"
repeat_region	11520. .11640	repeat_region	32639. .32958
repeat_region	/note="MST-INTERNAL repeat: matches 1520. .1650 of consensus"	repeat_region	/note="MIRB repeat: matches 1. .337 of consensus"
repeat_region	11681. .11934	repeat_region	33466. .33713
repeat_region	/note="L1M1-INTERNAL repeat: matches 964. .1212 of consensus"	repeat_region	33909. .33987
repeat_region	12067. .12284	repeat_region	/note="L2 repeat: matches 2639. .2748 of consensus"
repeat_region	/note="L1M1-INTERNAL repeat: matches 657. .889 of consensus"	repeat_region	34449. .34527
repeat_region	12513. .13673	repeat_region	/note="L1PA12 repeat: matches 5562. .5640 of consensus"
repeat_region	/note="L1M1-INTERNAL repeat: matches 109. .1231 of consensus"	repeat_region	34531. .34728
repeat_region	13674. .13986	repeat_region	/note="L1PA12 repeat: matches 1. .214 of consensus"
repeat_region	/note="AluSq repeat: matches 1. .313 of consensus"	repeat_region	34845. .35074
repeat_region	13987. .14017	repeat_region	/note="L1M1 repeat: matches 211. .453 of consensus"
repeat_region	/note="L1M1 repeat: matches 1. .109 of consensus"	repeat_region	35075. .35593
repeat_region	14019. .14109	repeat_region	/note="L1PA12 repeat: matches 5639. .6165 of consensus"
repeat_region	/note="L1M1-INTERNAL repeat: matches 405. .503 of consensus"	repeat_region	36282. .36570
repeat_region	14116. .14211	repeat_region	/note="AluSc repeat: matches 1. .289 of consensus"
repeat_region	/note="MST-INTERNAL repeat: matches 201. .293 of consensus"	repeat_region	36599. .37953
repeat_region	14247. .14469	repeat_region	/note="L1PA16 repeat: matches 4781. .6150 of consensus"
repeat_region	/note="L1M1D-internal repeat: matches 4. .225 of consensus"	repeat_region	39802. .39867
repeat_region	14472. .14810	repeat_region	/note="match: GSS AQJ180570"
repeat_region	/note="L1M1D repeat: matches 164. .505 of consensus"	repeat_region	40356. .40425
repeat_region	15542. .16687	repeat_region	/note="33 copies 2 mer ta 67% conserved"
repeat_region	/note="L2 repeat: matches 2602. .2745 of consensus"	repeat_region	40611. .40686
repeat_region	16724. .16792	repeat_region	/note="35 copies 2 mer tt 70% conserved"
repeat_region	/note="MIR repeat: matches 160. .233 of consensus"	repeat_region	41665. .42458
repeat_region	16875. .17286	repeat_region	/note="38 copies 2 mer tt 66% conserved"
repeat_region	/note="L2 repeat: matches 1774. .2234 of consensus"	repeat_region	42488. .42519
repeat_region	17606. .17975	repeat_region	/note="L1PA7 repeat: matches 5293. .6143 of consensus"
repeat_region	/note="THEIC repeat: matches 1. .371 of consensus"	repeat_region	42521. .43181
repeat_region	18633. .18864	repeat_region	/note="16 copies 2 mer at 94% conserved"
repeat_region	/note="MIR repeat: matches 9. .261 of consensus"	repeat_region	42521. .43181
repeat_region	19203. .19393	repeat_region	/note="L1M1 repeat: matches 5491. .6170 of consensus"
repeat_region	/note="MIR repeat: matches 339. .525 of consensus"	repeat_region	43796. .43949
repeat_region	19992. .20371	repeat_region	/note="match: 278716 chromosome 6 HindIII fragment"
repeat_region	/note="L2 repeat: matches 2318. .2749 of consensus"	repeat_region	44358. .44908
repeat_region	20461. .20585	repeat_region	/note="MIR repeat: matches 1. .452 of consensus"
repeat_region	/note="L2 repeat: matches 2541. .2667 of consensus"	repeat_region	44909. .44995
repeat_region	21143. .21217	repeat_region	/note="3 copies 29 mer 79% conserved"
repeat_region	/note="MIR repeat: matches 25. .108 of consensus"	repeat_region	44973. .45002
repeat_region	21555. .21850	repeat_region	/note="15 copies 2 mer ac 87% conserved"
repeat_region	/note="AluYo repeat: matches 5. .289 of consensus"	repeat_region	46196. .46355
repeat_region	22343. .22761	repeat_region	/note="MIR repeat: matches 47. .227 of consensus"
repeat_region	/note="L1M4C repeat: matches 1492. .1907 of consensus"	repeat_region	46558. .46916
repeat_region	22762. .23474	repeat_region	/note="L2 repeat: matches 975. .1346 of consensus"
repeat_region	/note="L1M4 repeat: matches 2732. .3507 of consensus"	repeat_region	49411. .49699
repeat_region	23475. .23618	repeat_region	/note="AluSc repeat: matches 1. .289 of consensus"

RESULT 15
 CELB0207 43584 bp DNA INV 27-MAY-1999
 LOCUS Caenorhabditis elegans cosmid B0207.
 DEFINITION
 ACCESSION U97196
 MID 91943799
 NID 097196.1 GI:1943799
 KEYWORDS
 SOURCE
 ORGANISM
 Caenorhabditis elegans.
 Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 43584)
 The C. elegans Genome Sequencing Consortium, Washington University
 Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,
 Hinxton, U.K., C.
 TITLE
 Genome sequence of the nematode C. elegans: a platform for
 investigating biology. The C. elegans Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 JOURNAL
 MEDLINE
 99069613
 Erratum: [[published erratum appears in Science 1999 Jan
 1;283(5398):35]]
 2 (bases 1 to 43584)
 Henkhaus, J. and Wohlmann, P.
 The sequence of C. elegans cosmid B0207
 JOURNAL
 REFERENCE
 3 (bases 1 to 43584)
 Waterston, R.
 Direct Submission
 Submitted (14-APR-1997)
 4 (bases 1 to 43584)
 Waterston, R.
 Direct Submission
 Submitted (27-MAY-1999) Department of Genetics, Washington
 University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
 COMMENT
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University,
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1RQ, England
 e-mail: r.w@genome.wustl.edu and j.s@sanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:
 all regions were double stranded, sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=
 30); an attempt was made to resolve all sequencing problems,
 such as compressions and repeats; all regions were covered by
 sequence from more than one m33 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' clone is Y76G2B, 3950 bp overlap; 3' clone is Y76G2A, 200 bp
 overlap. Actual start of this clone is at base position 1 of
 CELB0207; actual end is at 43584 of CELB0207

NOTES:

Coding sequences below are predicted from computer analysis, using
 the program GeneIndexer (P. Green and L. Hillier, ms in preparation).

FEATURES

Source
 1. 43584
 /organism="Caenorhabditis elegans"
 /strain="Bristol N2"

gene
 CDS

gene
 CDS

gene

/db_xref="taxon:6239"
 /clone="B0207"
 /chromosome="I"
 44..3900
 /gene="B0207.5"
 join(44..2693,2739..2873,3125..3204,3775..3900)
 /note="B0207.5"
 /note="coded for by C. elegans cDNA yk21b6.5; coded for by
 C. elegans cDNA yk5f10.3; coded for by C. elegans cDNA
 yk47d12.3; coded for by C. elegans cDNA yk130c11.3; coded
 for by C. elegans cDNA yk47d12.5; coded for by C. elegans
 cDNA yk130c11.5"
 /codon_start=1
 /protein_id="AAB52460.1"
 /db_xref="PID:g1943806"
 /db_xref="GI:1943806"
 /translation="MKSSTTTTTPAPPEYVDEEYPEVITLVKESRRPLKYSR
 PDDEDDDEVYREKPKSYITFESEFVYRGPRKIRKRPSTDSGEEDIREYRKIV
 EGEEDRGVLRKSHRKLVIRKQDEEPEEHEVRYVAKLVKSGNEKEFNDE
 ENVDDGKIIVRNRRKQYFERNDPDEEELVNRVAVRGAAPKIFRAHAPGR
 SAKPSITERNSEEFKIGSHEENHRKKNQKKNREDYFRRRRKPPEESELGR
 RKPRGESEEFERKSGRPDEONGETTVIKYSGSISDPTPLPDQESVAITK
 KIKFYGRKRDTPDDEENETIKLKRQLYIGARDKPTGNDNGELVGLITTKKSF
 IGPNGQKPTPGNDEDEPOFRKTSIDHEKODEVOLEFKKNTFRIVNGKVLRTTE
 DIPDDEGESVKITKKFRIQNGEITVDEDESRSRKISDEPEDDEDIKLF
 KRKFYKIVDKRVLITKEDVPQGGENGQIKTKKRYIIGNGEITKSSDDAL
 PSHDQGLVIGKNNRLYRKGEIPEDEENGESYVAKRKQLEKPOSEEEDEE
 FKLLKKKMGVIANPDSGSEIIKMSKKAVLYGSGMSDDODESEENMKR
 RFYVYGDGKREILREBEIPMGOLGTSKSSYSPDLGERTKRIYKIGRDGKOL
 OKMGDDESGDSQGTIVYKMKRKTIRYANSEKELIKENYPGDLSGDEGDKA
 IWSKRTKSTYIDGKNNKLVKEKRVGPEDEGQKLMGKTKRTKRYDGRKLM
 KTESDSDGNMMPKHNSTPNDGVPPGHSYVSKSDNNKRDGDSERGSDEER
 TRNGKTNKRDGRDGRNMNLSIIPRRSSNDGNDSRNDNRREKTRSSRARSS
 RKMEYKHDRVNRSSGSEELVPE"
 complement(4089..7564)
 /gene="B0207.6"
 complement(join(4089..4429,4757..4955,5113..5220,
 5338..5509,7482..7564))
 /gene="B0207.6"
 /note="weak similarity to ATPases"
 /codon_start=1
 /evidence-not_experimental
 /protein_id="AAB52461.1"
 /db_xref="PID:g1943807"
 /db_xref="GI:1943807"
 /translation="MOMLNKPKQLRVRYATSEKNGLGSTHTMGVLVIGAPGA
 GKSTGAGLIDITFISQTRPRTINLDPDNDMAVAPDYNTETMYNDVMDRLGPN
 GALKYCIETGANCNMLQKTEAHKKYLLIDCGQQLYISEBELKVIYFLKSGV
 RLCAHLADSLYCSDPKFISSALSTLAITVMTMPQVNCISKADLSESDGTLEEF
 SHLPVNRLLDLNLEPGLERKLRKLNALICGVISDFLVSEFPLAVENKESMRYIOM
 VDAANGFSLTEBOGIRELILNSKA"
 5897..7241
 /gene="B0207.4"
 join(5897..5961,6029..6103,6153..6261,6338..6705,
 6769..7042,7143..7241)
 /gene="B0207.4"
 /note="Contains similarity to Pfam domain: PF00069
 (pklnase). Score=295.1, E-value=2.8e-85, N=1; coded for by
 C. elegans cDNA CEST328f; coded for by C. elegans cDNA
 CEST94f; coded for by C. elegans cDNA CEST336f; coded for
 by C. elegans cDNA yk11a10.5"
 /codon_start=1
 /protein_id="AAB52459.1"
 /db_xref="PID:g1943805"
 /db_xref="GI:1943805"
 /translation="MYSIKRNMAIELINSITKFIYISDMENKPPYINLPEKTYNTPQ
 RGGKFTINDFEIGRPLGKRGFSYLAATKQGHVHALKVLFSQLISGVEHLERE
 IEIOSHLNHPNIIKLYTFWDKRIYLVLEAPGEGMKTQVTSKRSEPTAAAYME
 IADALSYCHRNKIVLHRIIDPENLLISGGLKIDGFSQVAPSNKRSEPTGTMDYLP
 PEMVNGADHSADVADMAIGVLYCYELVQKPPFEHDGSKTYAAIKAAFTYPSVKKG
 ARDLIGRLVYDPAKRCLEBQVKEHYVQGMMEKIRAKKQKTEKESLANH"
 11129..114089
 /gene="gpa-14"

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:59:37 ; Search time 255.05 Seconds

(without alignments)
22.562 Million cell updates/sec

Title: US-09-325-095-29

Perfect score: 23

Sequence: 1 RAARTGCCRRARTCDARTGRGA 23

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N.Geneseq.36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73.9	9789	1	T41852	CDNA encoding plas
2	72.9	6831	1	T80200	Phosphatidyl inosi
3	72.2	38	1	Q59013	Primer for p110 CD
4	72.2	38	1	Q57019	Ptdins 3-kinase pr
5	72.2	1610	1	X04612	Genomic sequence e
6	69.6	1131	1	Q48032	Flnd DNA. Vaccines
7	69.6	641	1	V75788	Staphylococcus aur
8	68.7	40875	1	T80043	Insert from cosmid
9	67.8	2540	1	Q29627	Hepatitis C virus
10	67.8	2540	1	Q43888	NANB hepatitis vir
11	67.8	2401	1	Q48539	Bacillus subtilis
12	67.8	381	1	Q57522	Human P17R-f CDNA.
13	67.8	393	1	Q57523	Human P17R-f CDNA.
14	67.8	24	1	Q59023	Degenerate primer
15	67.8	2540	1	Q63752	NANBH genomic fra
16	67.8	9502	1	Q74770	Hepatitis C virus
17	67.8	1835	1	T29773	Maize adenylosucci
18	67.8	1835	1	T32829	Maize adenylosucci
19	67.8	3141	1	T71746	Fltrobacter succin
20	67.8	3387	1	V16533	CDNA encoding mam
21	67.8	5220	1	V31340	Human phosphatidyl
22	67.8	3549	1	X13796	Enterococcus faeca
23	67.8	8921	1	X13005	Enterococcus faeca
24	67.8	537	1	X13917	Enterococcus faeca
25	67.8	5220	1	X15932	CDNA sequence enco
26	67.8	1835	1	X25765	Maize adenylosucci
27	67.8	1946	1	N70526	Sequence of CDNA 1
28	67.0	110000	1	T58840	Continuation (4 of
29	67.0	997	1	V15021	652 strain leukoto
30	67.0	992	1	V15030	652 strain leukoto
31	67.0	5285	1	T80199	Phosphatidyl inosi
32	67.0	3140	1	V52291	Streptococcus pneu
33	67.0	5061	1	V42920	CDNA encoding a hu
34	67.0	32768	1	X13060	Enterococcus faeca
35	67.0	995	1	X13572	Enterococcus faeca
36	67.0	529	1	X30828	Streptococcus pneu
37	67.0	992	1	X28835	A.actinomycetemcom
38	66.1	2031	1	V59655	Human secreted pro
39	65.2	2111	1	O63889	Sequence coding de
40	65.2	2774	1	O66579	Sorbiol dehydroge
41	65.2	2664	1	O78644	E.coli pyruvate de
42	65.2	8041	1	T10105	Adherence conferr
43	65.2	3707	1	O99277	Human neurotrophic

ALIGNMENTS

C	44	15	65.2	3194	1	T00689	Human trkB recepto
	45	15	65.2	3515	1	X20252	Borrelia burgdorfe
ALIGNMENTS							
RESULT	1						
ID	T41852						
AC	T41852						
DT	20-FEB-1997						
DE	CDNA encoding plasmodium falciparum erythrocyte membrane protein.						
KW	Plasmodium falciparum; erythrocyte membrane protein; malaria; detection; identification; treatment; prevention; parasite; ss.						
OS	Plasmodium falciparum MC type.						
FM	Key						
FT	Location/Qualifiers						
FT	cds						
FT	326..9497						
FT	/*tag= a						
FT	/product= Erythrocyte membrane protein						
FT	518..520						
FT	/*tag= b						
FT	/transl_except= GNA encodes Tyrosine						
FT	656..658						
FT	/*tag= c						
FT	/transl_except= ATT encodes Leucine						
FT	2909..2911						
FT	/*tag= d						
FT	/transl_except= AAC encodes Aspartic acid						
FT	3461..3463						
FT	/*tag= e						
FT	/transl_except= GAA encodes Glutamine						
FT	5546..5548						
FT	/*tag= f						
FT	/transl_except= CCG encodes Arginine						
FT	6254..6256						
FT	/*tag= g						
FT	/transl_except= AAT encodes Lysine						
FT	6257..6259						
FT	/*tag= h						
FT	/transl_except= ATA encodes Tyrosine						
FT	6263..6265						
FT	/*tag= i						
FT	/transl_except= AAC encodes Lysine						
FT	6269..6271						
FT	/*tag= j						
FT	/transl_except= TTC encodes Isoleucine						
FT	6272..6274						
FT	/*tag= k						
FT	/transl_except= ATA encodes Histidine						
FT	6275..6277						
FT	/*tag= l						
FT	/transl_except= ATT encodes Asparagine						
FT	6278..6280						
FT	/*tag= m						
FT	/transl_except= GGA encodes Tryptophan						
FT	7754..8478						
FT	/*tag= n						
PN	intron						
PD	MO9633736-A1.						
PD	31-OCT-1996.						
PR	26-APR-1996; U05798.						
PR	27-APR-1995; US-430908.						
PA	(AFRY-) AFFYMAX TECHNOLOGIES NV.						
PI	Baruch DI, Howard RJ, Pasloske BL;						
DR	WPI: 96-497376/49.						
DR	P-FCDB: W00384.						
PT	New Plasmodium falciparum erythrocyte membrane proteins - used to						
PT	develop products for the diagnosis, treatment or prevention of						
PT	malaria parasite infections						
PS	Disclosure: Figure 12: 149pp; English.						
CC	A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte						
CC	membrane protein 1 (PfEMP1) or active fragments or analogues of that						
CC	protein can be used in the treatment or prevention of symptoms of a						

CC malaria parasite infection. The polypeptides can inhibit, block or
 CC reverse the sequestration of erythrocytes in patients suffering from
 CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
 CC probes and primers to identify a Plasmodium falciparum parasite, the
 CC primers used to generate characteristic amplification patterns from
 CC different P. falciparum strains. Antibodies specifically
 CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
 CC used in diagnosis of malaria infection. This sequence encodes the
 CC PfEMP1 protein of the MC type of Plasmodium falciparum. An
 CC alternative, truncated version of the coding sequence (a cDNA clone)
 CC is given in T41853.
 SQ Sequence 9789 BP; 4061 A; 1393 C; 1837 G; 2498 T;

Query Match 73.9%; Score 17; DB 1; Length 9789;
 Best Local Similarity 65.2%; Pred. No. 25;
 Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGRGA 23
 :|||:|||||:|||||:|||||:
 DB 6232 AACTGCAAAATCATCATGTGAAA 6254

RESULT 2
 ID 180200 standard; cDNA; 6831 BP.
 AC 180200;
 DT 22-JUN-1998 (first entry)
 DE Phosphatidyl inositol 3-kinase: signal transduction; cell cycle;
 KM Phosphatidyl inositol 3-kinase; cell proliferation; cancer;
 KW antagonist; inflammatory joint disease; cell proliferation; cancer;
 KM psoriasis; restenosis; atherosclerosis; therapy; diagnosis; cpk; ss.
 OS Drosophila melanogaster.
 FT Key Location/Qualifiers
 FT CDS 148..5778
 FT /tag- a
 PN WO9731650-A1.
 PD 04-SEP-1997.
 PF 12-FEB-1997; U02193.
 PR 29-FEB-1996; US-609049.
 PA (REGC) UNIV CALIFORNIA.
 PI Chen Y, Moiz L, Williams LT;
 PI WPI: 97-448442/41.
 DR P-PSDB; W38757.
 PT New isolated phosphatidyl inositol-3 kinase polypeptide - used to
 PT develop products for diagnosis and therapy, particularly for
 PT proliferative disorders, e.g. inflammatory joint diseases, or cancer
 PS Example 1; Fig 9; 77pp; English.
 CC This cDNA sequence codes for cpk (see W38757), a Drosophila
 CC polypeptide that belongs to a novel class of phosphatidyl inositol
 CC 3-kinases that contain a C2 domain, are capable of phosphorylating
 CC a D3 hydroxyl of an inositol ring in phosphatidyl inositol and
 CC phosphatidyl inositol 4-phosphate, but not in phosphatidyl inositol
 CC 4,5-bisphosphate, and which are involved in cell signalling cascades
 CC that control e.g. cell cycle progression and intracellular protein
 CC sorting. Short fragments of cpk cDNA were obtained from a
 CC Drosophila cDNA library by PCR (see T80195-96). These short
 CC fragments were used to screen the cDNA library to obtain larger
 CC fragments, and missing 5' ends were obtained by RACE. A
 CC recombinant host cell, transfected with a vector comprising a cpk
 CC nucleic acid is claimed. Novel phosphatidyl inositol 3-kinases
 CC can be used to screen for agonists/antagonists of activity and
 CC in a claimed method of treating a disorder caused by dysregulation
 CC of a growth factor activation signalling cascade. Antagonists
 CC may reduce Ras activation allowing treatment of proliferative
 CC disorders such as atherosclerosis, inflammatory joint disease,
 CC psoriasis, restenosis following angioplasty, and cancer. 1648 T;
 SQ Sequence 6831 BP; 1829 A; 1683 C; 1669 G; 1648 T;

Query Match 73.9%; Score 17; DB 1; Length 6831;
 Best Local Similarity 65.2%; Pred. No. 24;
 Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGRGA 23
 :|||:|||||:|||||:|||||:
 DB 4529 AACTGCAAAATCATCATGTGAAA 4507

RESULT 3
 ID 059013 standard; DNA; 38 BP.
 AC 059013;
 DT 12-APR-1994 (first entry)
 DE Primer for p110 cDNA.
 KM Phosphoinositide kinase; PI; p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification.
 OS Synthetic.
 PN WO9321328-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 PI WPI: 93-351736/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 1; Page 38; 146pp; English.
 CC An SGBAF-1 cell line was established by transfection of bovine
 CC adrenal cortex zona fasciculata cells with pSVneo. Total RNA was
 CC isolated from the SGBAF-1 and a cDNA library constructed. The
 CC unamplified library was plated on E. coli K12 PLF-F' and screened
 CC with the labelled primer shown and the primer of 059012. Hybridising
 CC clones were sequenced.
 CC See also 051155-6, 059012-23 and 057522-3.
 SQ Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 72.2%; Score 16.6; DB 1; Length 38;
 Best Local Similarity 73.9%; Pred. No. 15;
 Matches 17; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAARTGCCRAARTCDATRTGRGA 23
 :|||:|||||:|||||:|||||:
 DB 38 AARTGCCAAARTCATRTGRGA 16

RESULT 4
 ID 057019 standard; DNA; 38 BP.
 AC 057019;
 DT 31-AUG-1994 (first entry)
 DE Ptdins 3-kinase probe #2.
 KW 110 kD catalytic subunit; phosphatidyl inositol 3-kinase;
 KW transformation; Schistosomaccharomyces pombe; nmt promoter; thiamine;
 KW Ptdins 3-kinase; assay; detection; cell growth; regulation; cancer;
 KW blood vessel plaques; ss.
 OS Synthetic.
 PN WO9403609-A.
 PD 17-FEB-1994.
 PF 05-AUG-1993; G01651.
 PR 05-AUG-1992; GB-016654.
 PA (IMCR) IMPERIAL CANCER RES TECHNOLOGY.
 PI Goode NT, Nurse PM, Parker PJ, Waterfield MD;
 PI WPI: 94-065697/08.
 PT Eukaryotic cells transformed with mammalian phospholipid or
 PT protein kinase DNA - useful in assays for compounds involved in
 PT cell growth regulation and for treating cancers
 PS Example 1; Page 14; 71pp; English.
 CC The sequences given in 057018-19 are probes which were used in the
 CC isolation of fusion sequences comprising the phosphatidyl inositol
 CC (Ptdins) 3-kinase under the regulatory control of the nmt promoter.
 CC These sequences were transformed into Schistosomaccharomyces pombe cells

CC in an embodiment of the invention. In the presence of thiamine the
 CC promoter is inactive and the cells carrying the PKC plasmids grow as
 CC the parental strain. In the absence of thiamine the nmt promoter
 CC functions and the PKC is induced. PKC activity is substantially
 CC increased under these conditions. Cells containing constructs such as
 CC this, are useful in assays for detecting compounds involved in cell
 CC growth regulation. It is also used as the basis for detecting
 CC compounds for treating cancers and the formation of blood vessel
 CC plaques.
 SQ Sequence 38 BP; 7 A; 6 C; 8 G; 11 T;

Query Match 72.2%; Score 16.6; DB 1; Length 38;
 Best Local Similarity 73.9%; Pred. NO. 15;
 Matches 17; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 1 RAARTGCCRAARTCDATRTGRAA 23
 :||:|||||:|||||
 DB 38 RAARTGCCRAARTCDATRTGRAA 16

RESULT 5
 X04612
 ID X04612 standard; cDNA; 1610 BP.
 AC X04612.
 DT 12-APR-1999 (first entry)
 DE Genomic sequence encoding the GA4 homologue GA4H1.
 KW Gibberellin 4; GA4; beta-hydroxylase; GA4H; GA4H1; GA4H2;
 KM plant growth hormone; seed germination; stem elongation; flowering;
 KM fruiting; stem growth; ss.
 OS Arabidopsis thaliana.
 FH Key
 FT primer_bind
 32..51
 Location/Qualifiers
 FT /tag= a
 /note= "binding site for primer GA-P15 (X04624)"
 43..62
 FT /tag= b
 /note= "binding site for primer GA-P14S (X04628)"
 86..1562
 FT /tag= c
 /product= GA4H1
 FT /transl_except= (pos: 554..965, aa: Trp)
 /note= "the sequence contains 1 intron, and encodes
 the full length protein W88003; the
 translation exception occurs at the point where
 the exons join"
 86..555
 FT /tag= d
 /number= 1
 /note= "this exon encodes W88000"
 484..503
 FT /tag= e
 /note= "binding site for primer GA-P2 (X04617)"
 503..520
 FT /tag= f
 /note= "binding site for primer GA-913 (X04621)"
 556..964
 FT /tag= g
 940..959
 FT /tag= h
 /note= "binding site for antisense primer GA-P6 (X04618)"
 965..1559
 FT /tag= i
 /number= 2
 /note= "this exon encodes W88002"
 1111..1130
 FT /tag= j
 /note= "binding site for antisense primer GA-P17
 (X04622)"
 1567..1584
 FT /tag= k
 /note= "binding site for antisense primer GA-P1X
 (X04623)"

FT primer_bind 1583..1597
 FT /tag= l
 /note= "binding site for antisense primer GA-P16
 (X04625)"
 PN MO9859057-A1.
 PD 30-DEC-1998.
 PF 24-JUN-1998; U13044.
 PR 24-JUN-1997; US-050615.
 PA (CHIA) CHIANG H.
 PA (GEO) GEN HOSPITAL CORP.
 PA (GOOD) GOODMAN H M.
 PA (NGUY) NGUYEN L V.
 PI Chlang H, Goodman HM, Nguyen LV;
 PI WPI; 99-105626/09.
 DR P-PSDB; W88000, W88002, W88003.
 DE New isolated Gibberellin 4 homologues - derived from Arabidopsis
 PT plants, used to develop products for altering stem growth, e.g. for
 PT enhancing stem elongation, flowering and fruiting
 PS Claim 4; Fig 6A-B; 104pp; English.
 CC The present sequence represents the genomic sequence of a gibberellin
 CC 4 (GA4) homologue, GA4H1. The GA4H proteins (GA4H1 and GA4H2) have
 CC similar functions to GA4. GA4H is believed to be a member
 CC of the enzyme family involved in the biosynthesis of the gibberellin
 CC family of plant growth hormones that promote various growth and
 CC developmental processes in higher plants, such as seed germination,
 CC stem elongation, flowering and fruiting. GA4 is a beta-hydroxylase, and
 CC the homologues may also have 3-beta-hydroxylase activity.
 CC 3-beta-hydroxylase is critical for controlling stem growth. The GA4H
 CC may be applied to crops to enhance and facilitate stem elongation,
 CC flowering and fruiting. Alternatively, the DNA encoding GA4H may be
 CC genetically inserted into the plant host to produce a similar effect.
 SQ Sequence 1610 BP; 459 A; 357 C; 325 G; 469 T;

Query Match 72.2%; Score 16.6; DB 1; Length 1610;
 Best Local Similarity 60.9%; Pred. NO. 29;
 Matches 14; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

OY 1 RAARTGCCRAARTCDATRTGRAA 23
 :||:|||||:|||||
 DB 845 RAARTGCCRAARTCDATRTGRAA 867

RESULT 6
 Q48032/c
 ID Q48032 standard; DNA; 1131 BP.
 AC Q48032.
 DT 08-FEB-1994 (first entry)
 DE FimD DNA.
 KW N-terminal; minor; fimbrial subunit; B. pertussis; fimbriae; FimD;
 KW vaccine; whooping cough; immune response; B. pertussis;
 KW B. bronchiseptica; ss.
 OS Bordetella pertussis.
 FH Key
 FT signal_peptide 1..111
 FT /tag= a
 /note= "binding site for primer GA-P15 (X04624)"
 112..1131
 FT /tag= b
 /note= "binding site for primer GA-P14S (X04628)"
 1132..1562
 FT /tag= c
 /product= GA4H1
 FT /transl_except= (pos: 554..965, aa: Trp)
 /note= "the sequence contains 1 intron, and encodes
 the full length protein W88003; the
 translation exception occurs at the point where
 the exons join"
 86..555
 FT /tag= d
 /number= 1
 /note= "this exon encodes W88000"
 484..503
 FT /tag= e
 /note= "binding site for primer GA-P2 (X04617)"
 503..520
 FT /tag= f
 /note= "binding site for primer GA-913 (X04621)"
 556..964
 FT /tag= g
 940..959
 FT /tag= h
 /note= "binding site for antisense primer GA-P6 (X04618)"
 965..1559
 FT /tag= i
 /number= 2
 /note= "this exon encodes W88002"
 1111..1130
 FT /tag= j
 /note= "binding site for antisense primer GA-P17
 (X04622)"
 1567..1584
 FT /tag= k
 /note= "binding site for antisense primer GA-P1X
 (X04623)"

CC The polypeptide it encodes may be used in a system for detecting
CC NANB hepatitis. This method is highly specific and sensitive, and
CC can detect NANB hepatitis virus which could not be detected by
CC conventional methods.

SQ Sequence 2540 BP; 477 A; 773 C; 741 G; 549 T;

Query Match 67.8%; Score 15.6; DB 1; Length 2540;

Best Local Similarity 65.0%; Pred. No. 97;

Matches 13; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 3 ARTGCCAATCDAATRTGAA 22
1:|||||:|:|:|:|:|:|
DB 2193 ARTGCCAAGCCATATGGA 2174

RESULT 11

ID 048539 standard; cDNA: 2401 BP.

AC 048539;

DT 21-FEB-1994 (first entry)

DE Bacillus subtilis inositol dehydrogenase gene.

KW Recombinant protein production; inositol dehydrogenase; ss.

OS Bacillus subtilis.

FH Key

FT cds

FT location/Qualifiers

FT /tag=2

FT /product= inositol_dehydrogenase

PN J05192163-A.

PD 03-AUG-1993.

PF 30-SEP-1991: 252073.

PR 30-SEP-1991: JP-252073.

PA (SUNR) SUNTORY LTD.

DR WPI: 93-277476/35.

P-PSDB: R40918.

PT Inositol dehydrogenase gene - isolated from e.g. bacteria

PI belonging to Bacillus subtilis, in large amt.

PS Claim 4: Page 8-10; 18pp; Japanese.

CC The Bacillus subtilis inositol dehydrogenase gene has been cloned

CC and sequenced. Microorganisms transformed by the coding sequence can

CC be cultured to produce large amounts of recombinant inositol

CC dehydrogenase.

SQ Sequence 2401 BP; 657 A; 517 C; 609 G; 618 T;

Query Match 67.8%; Score 15.6; DB 1; Length 2401;

Best Local Similarity 65.0%; Pred. No. 96;

Matches 13; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 3 ARTGCCAATCDAATRTGAA 22
1:|||||:|:|:|:|:|:|

DB 1998 ARTGCTGAATCATATGGA 2017

RESULT 12

ID 057522/c

AC 057522 standard; cDNA: 381 BP.

DT 12-APR-1994 (first entry)

DE Human P1TR-c cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ss.

PN WO9321328-A.

PD 28-OCT-1993.

PF 13-APR-1993: G00761.

PR 13-APR-1993: GB-008135.

PA (LUDWIG) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

P-PSDB: R46552.

Query Match 67.8%; Score 15.6; DB 1; Length 393;

Best Local Similarity 72.2%; Pred. No. 70;

Matches 13; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 6 GCCRAATCDAATRTGAA 23

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PT activity, useful for controlling cell proliferation

PS Disclosure; Fig 20; 14pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC glucose levels can be controlled using the kinase.

CC See also 051155-6, 059012-23 and 057523.

CC Sequence 381 BP; 116 A; 78 C; 85 G; 102 T;

SQ Sequence 381 BP; 116 A; 78 C; 85 G; 102 T;

Query Match 67.8%; Score 15.6; DB 1; Length 381;

Best Local Similarity 72.2%; Pred. No. 70;

Matches 13; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 6 GCCRAATCDAATRTGAA 23
1:|||||:|:|:|:|:|:|

DB 378 GCCGAATCGATGTGGA 361

RESULT 13

ID 057523/c

AC 057523 standard; cDNA: 393 BP.

DT 12-APR-1994 (first entry)

DE Human P1TR-f cDNA.

KW Phosphoinositide kinase; PI; p85 subunit; screening; agonist;

KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;

KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;

KW ss.

PN WO9321328-A.

PD 28-OCT-1993.

PF 13-APR-1993: G00761.

PR 13-APR-1993: GB-008135.

PA (LUDWIG) LUDWIG INST CANCER RES.

PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;

PI ParkerPJ, Volinia S, Waterfield MD;

DR WPI: 93-351738/44.

P-PSDB: R46553.

PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase

PT activity, useful for controlling cell proliferation

PS Disclosure; Fig 21; 14pp; English.

CC Southern blot analysis was performed using a bovine cDNA probe contg.

CC a fragment of a PI3-kinase-encoding sequence and human cDNA isolated

CC from a cDNA library constructed from mRNA isolated from the human

CC cell line K562. Positive clones were sequenced to give the human

CC PI3 kinase p110 sequence shown. This sequence has 95 percent

CC homology with the bovine sequence. The domain encoding residues 19-

CC 100 of human p110 is sufficient to encode the kinase which will

CC associate with the p85 kinase subunit. The gene may be used to

CC provide a protein with PI3 kinase activity and is useful for

CC screening for (ant)agonists of PI3 kinase activity which could be

CC useful for stimulation or inhibition of cell proliferation and hence

CC glucose levels can be controlled using the kinase.

CC See also 051155-6, 059012-23 and 057522.

CC Sequence 393 BP; 87 A; 103 C; 124 G; 79 T;

SQ Sequence 393 BP; 87 A; 103 C; 124 G; 79 T;

Db 393 CCCGAGTCGATGTGGA 376

RESULT 14

059023 standard: DNA: 24 BP.

AC 059023: 12-APR-1994 (first entry)
 DE Degenerate primer for p110 cDNA.
 KW Phosphoinositide kinase; PI: p85 subunit; screening; agonist;
 KW antagonist; cell proliferation; inhibition; prophylaxis; therapy;
 KW platelets; neutrophil activity; 3-phosphorylated phosphoinositides;
 KW ss; amplification.
 OS Synthetic.
 FH Key
 FT misc_difference 7 Location/Qualifiers
 FT /tag= a
 FT /note= "inosine"
 FN WO9321228-A.
 PD 28-OCT-1993.
 PF 13-APR-1993; G00761.
 PR 13-APR-1992; GB-008135.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Dhand R, Fry MJ, Gout I, Hiles ID, Otsu M, Panayotou G;
 PI Parker PJ, Volinia S, Waterfield MD;
 DR WPI: 93-351738/44.
 PT Recombinant polypeptide(s) - with phosphoinositide-3 kinase
 PT activity, useful for controlling cell proliferation
 PS Example 2; Page 53; 146pp; English.
 CC Two novel cDNAs related to p110 have been cloned. Degenerate
 CC primers were designed to conserved sequences between human p110 and the
 CC related yeast gene VPS34. These were used in RT-PCR using mRNA from the
 CC human cell lines MOLT4 and U937. Two novel cDNAs, P1TR-C and P1TR-F,
 CC related to p110, were isolated.
 CC See also Q51155-6, Q59012-22 and Q57522-3.
 SQ Sequence 24 BP; 7 A; 3 C; 2 G; 4 T;

Query Match

Best Local Similarity 67.8%; Score 15.6; DB 1; Length 24;
 Best Local Similarity 78.3%; Pred. No. 43;
 Matches 18; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 RAATGCCRAARTCDATRTGAA 23

Db 2 AATGCCRAARTCDATRTGAA 24

RESULT 15

063752/c ID 063752 standard: cDNA to genomic RNA: 2540 BP.

AC 063752: 30-JAN-1995 (first entry)
 DE NANBHV genomic fragment.
 KW Polymerase chain reaction; PCR; primer; amplify; detection; NANBHV;
 KW non-A, non-B hepatitis virus; 5'-terminal region; core protein; ss.
 OS Synthetic.
 PN J06125777-A.
 PD 10-MAY-1994.
 PF 20-JUN-1991; 247120.
 PR 20-JUN-1991; JP-247120.
 PA (NAKA/) NAKAMURA T.
 DR WPI: 94-187937/23.
 PT Oligonucleotide primer pairs specific for non-A, non-B hepatitis
 PT virus - used for high sensitivity detection of non-A non-B (NANB)
 PT hepatitis virus
 PS Disclosure: Page 22-23; 25pp; Japanese.
 CC The sequences given in Q63752-53 represent fragments of the non-A,
 CC non-B hepatitis virus (NANBHV) genome. These fragments were amplified
 CC using the primers given in Q63732-51. These primers were used in the
 CC detection of NANBH. The primers are based on the 5'-terminal region and
 CC the core protein coding region. The method allows highly sensitive
 CC detection of NANBH.

SQ Sequence 2540 BP; 477 A; 777 C; 737 G; 549 T;

Query Match

Best Local Similarity 67.8%; Score 15.6; DB 1; Length 2540;
 Best Local Similarity 65.0%; Pred. No. 97;
 Matches 13; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 3 AATGCCRAARTCDATRTGAA 22

Db 2193 AATGCCAAAGCCTATATGGA 2174

Search completed: September 13, 1999, 15:59:38
 Job time: 4805 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 15:45:57 ; Search time 1694.61 Seconds
(without alignments)

26.772 Million cell updates/sec

Title: US-09-325-095-29

Perfect score: 23
Sequence: 1 RAAATGCCRAATCDAITRTGAA 23

Scoring table: IDENTITY_NUC
2546578 segs, 98626752 residues

Database :

EST:*
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.2	74.8	740	45	AU004743	AU004743 AU004743
2	17	73.9	435	27	M99426	M99426 TGEEST201b1
3	17	73.9	481	29	AA153179	AA153179 mm31a09.r
4	17	73.9	269	30	AA240189	AA240189 mw27c04.r
5	17	73.9	647	51	A1728321	A1728321 BNGH1104
6	16.6	72.2	376	35	C45088	C45088 C45088 Yuj1
7	16.6	72.2	267	36	C67578	C67578 C67578 Yuj1
8	16.6	72.2	756	45	AU002985	AU002985 AU002985
9	16.6	72.2	658	45	AU006090	AU006090 AU006090
10	16.6	72.2	423	53	HSK005641	HSK005641
11	16.4	71.3	373	36	C69479	C69479 C69479 Yuj1
12	16.2	70.4	319	20	T02063	T02063 WEST02784 E
13	16.2	70.4	362	20	Z21659	Z21659 HSDR32E12 S
14	16.2	70.4	283	23	D60457	D60457 HUM110E12A
15	16.2	70.4	360	36	C61779	C61779 C61779 Yuj1
16	16.2	70.4	360	36	C67000	C67000 C67000 Yuj1
17	16.2	70.4	657	46	A1455444	A1455444 LD24279.3
18	16.2	70.4	299	50	AV045367	AV045367 AV045367
19	16	69.6	192	20	Z40168	Z40168 HSC1UD042 n
20	16	69.6	479	25	N76563	N76563 YV40803.r1
21	16	69.6	498	28	AA073214	AA073214 mm93c03.r
22	16	69.6	360	35	C42845	C42845 C42845 Yuj1
23	16	69.6	440	36	AA629895	AA629895 ad45b11.s
24	16	69.6	285	40	AA912476	AA912476 O196e07.s
25	16	69.6	330	40	AA912872	AA912872 O127a03.s
26	16	69.6	224	40	AA922454	AA922454 om63b11.s
27	15.8	68.7	482	27	AA001587	AA001587 z649f01.s
28	15.8	68.7	420	29	AA191170	AA191170 zq45a02.r
29	15.8	68.7	528	35	C23190	C23190 C23190 Japa
30	15.8	68.7	300	35	C53168	C53168 C53168 Yuj1
31	15.8	68.7	600	36	AA601777	AA601777 NCP4E3T7
32	15.8	68.7	303	38	AA784401	AA784401 d5d08a1.r
33	15.8	68.7	599	38	AA786082	AA786082 j5b02a1.r
34	15.8	68.7	526	38	AA786117	AA786117 j5b05a1.r
35	15.8	68.7	538	40	AA933666	AA933666 om59b03.s
36	15.8	68.7	335	40	AA965589	AA965589 k5a12a1.r
37	15.8	68.7	511	40	AA966375	AA966375 w4e02a1.r
38	15.8	68.7	479	40	AA966448	AA966448 w5b02a1.r
39	15.8	68.7	411	43	A1209373	A1209373 a0c0a1.r
40	15.8	68.7	428	48	A1572092	A1572092 tr73e06.x
41	15.8	68.7	768	49	A1620707	A1620707 tu96a12.x
42	15.8	68.7	588	50	A1675336	A1675336 w98b06.x
43	15.6	67.8	346	20	D24320	D24320 RICR1727A R
44	15.6	67.8	399	20	T10141	T10141 seg1375 b4H
45	15.6	67.8	236	24	N20317	N20317 yx44a11.s1

ALIGNMENTS

RESULT 1
AU004743
LOCUS AU004743 740 bp mRNA
DEFINITION AU004743 Bombyx mori p50(Daizo) Bombyx mori cDNA clone ws20797,
mRNA sequence.
ACCESSION AU004743
NID 94162114
VERSION AU004743.1 GI:4162114

```

KEYWORDS      EST.
SOURCE        domestic silkworm.
ORGANISM      Bombyx mori
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
              Ditrysia; Bombycoidea; Bombycidae; Bombyx.
REFERENCE     1 (bases 1 to 740)
AUTHORS       Mita,K., Morimyo,M., Shimada,T., Okano,K. and Maeda,S.
TITLE         Establishment of cDNA database of Bombyx mori
              Unpublished (1999).
JOURNAL       On Jan 14, 1998 this sequence version replaced gi:1798193.
COMMENT

FEATURES
  source
    Contact: Mita K
    Genome Research Group
    National Institute of Radiological Sciences
    Anagawa 4-9-1, Inage, Chiba 263-8555, Japan
    Email: kmिता@nirs.go.jp
    PROJECT = 'CREST project by JST'.
    Location/Qualifiers
      1..740
        /organism="Bombyx mori"
        /strain="p50(Daizo)"
        /db_xref="taxon:7091"
        /map="16"
        /cclone="ws20797"
        /cclone_1lb="Bombyx mori p50(Daizo)"
BASE COUNT    225 a      163 c      154 g      198 t
ORIGIN

Query Match      74.8%; Score 17.2; DB 45; Length 740;
Best Local Similarity 63.6%; Pred. No. 1.8e+02;
Matches 14; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

OY      1 RAARTGCCAARTCDATRTGRA 22
        ::|||::|||::|||::|||
Db      425 AAAGTCCGCAAGTCAACATGAA 446

RESULT 2
W99426      435 bp      mRNA      EST      11-SEP-1997
LOCUS       TG8Z201b12.r1.Tgmk49.Tachyzoite cDNA Toxoplasma gondii cDNA clone
DEFINITION  tg8z201b12.r1 5', mRNA sequence.
ACCESSION   W99426
VERSION     W99426.1 GI:1435312
KEYWORDS    EST.
SOURCE      Toxoplasma gondii.
            Toxoplasma gondii
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Sarcocystidae; Toxoplasma.
REFERENCE   1 (bases 1 to 435)
AUTHORS     Hohl,A., Manger,I., Marra,M., Sibley,L.D., Ajioke,J.A.,
            Asietie,M.A., Dietrich,N., Dubuque,T., Hillier,L., Kucaba,T.,
            Wan,K.L., Waterston,R.H. and Boothroyd,J.
            WashU-Merck-Stanford-NIH Toxoplasma EST project
            Unpublished (1996)
            On Jan 25, 1995 this sequence version replaced gi:637759.
COMMENT

TITLE       Contact: Marra M
JOURNAL     WashU-Merck EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel.: 314 286 1800
            Fax: 314 286 1810
            Email: toxo@watson.wustl.edu
            Clones are available from Genome Systems (genome@MO.NET); the
            library is available from John Boothroyd
            (jbooth@pepper.stanford.edu)
            Seq primer: T3
            *High quality sequence stop: 341.
            Location/Qualifiers
FEATURES

```

```

source
1. .435
/organism="Toxoplasma gondii"
/strain="ME49, Clone PDS"
/db_xref="taxon:5811
/map="10"
/clone="tgzz01b12.r1"
/clone_11b="TgME49 Tachyzoite CDNA"
/label="X11-Blue MRF"
/note="Vector: lambda ZAP; Site_1: EcoRI; Site_2: XhoI;
Toxoplasma parasites were grown in human foreskin
fibroblasts cultures in vitro. The library was
constructed by A. Hehl and I. Manger, Stanford University
cDNAs were synthesized from polyA mRNAs by oligo d(T)
priming and directionally cloned into the EcoRI and XhoI
sites of the lambda zap vector using the ZAP-cDNA
synthesis kit (Stratagene). Warning: the library contains a
small percentage of human cDNAs derived from the human
host cells."

BASE COUNT      118 a      88 c      104 g      123 t      2 others
ORIGIN

Query Match      73.9% Score 17; DB 27; Length 435;
Best Local Similarity 65.2%; Fred. NO. 2.3e+02;
Matches 15; Conservative 6; Mismatches 2; Indels 0; Gaps 0

QY      1 RAAATGCCCAARTCDATRTGRGA 23
        ||:||||:||||:| ||:|
Db      51 GAAGTCCGAGAGTCTTTTGAAA 29

RESULT 3
AA153179      481 bp      mRNA      EST      11-DEC-1996
LOCUS      ma13a09.r1 Beddington mouse embryonic region Mus musculus cDNA
DEFINITION      clone IMAGE:539512 5', mRNA sequence.
ACCESSION      AA153179
NID      g1124893
KEYWORDS      AA153179.1 GI:1724893
SOURCE      EST.
ORGANISM      house mouse.
            Mus musculus.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 481)
REFERENCE      Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
AUTHORS      Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
TITLE      The WashU-HM Mouse EST Project
JOURNAL      Unpublished (1996)
COMMENT      On May 17, 1996 this sequence version replaced g1:1182837.

Contact: Marra M/Mouse EST Project
WashU-HM Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:326448
Seq primer: -40m13 ET
High quality sequence stop: 393.
Location/Qualifiers
1. .481
/organism="Mus musculus"
/strain="C57BL/6 x DBA"
/db_xref="taxon:10090"
/map="21q22.1-3"
/clone="IMAGE:539512"
FEATURES
source

```


VERSION AU006090.1 GI:4163474
 EST.
 KEYWORDS
 SOURCE domestic silkworm.
 ORGANISM Bombyx mori
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia; Bombycoidea; Bombycidae; Bombyx.
 1 (bases 1 to 658)
 Mita, K., Moriyasu, M., Shimada, T., Okano, K. and Maeda, S.
 Establishment of cDNA database of Bombyx mori
 Unpublished (1999)
 JOURNAL
 COMMENT On Jan 17, 1998 this sequence version replaced gi:1900538.

FEATURES
 source
 1. 658
 /organism="Bombyx mori"
 /strain="p50(Daizo)"
 /db_xref="taxon:7091"
 /clone="v40467"
 /clone_lib="Bombyx mori p50(Daizo)"
 BASE COUNT 232 a 91 c 100 g 235 t
 ORIGIN

Query Match 72.2%; Score 16.6; DB 45; Length 658;
 Best Local Similarity 66.7%; Pred. No. 3.4e+02;
 Matches 14; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY 3 AATGCCCAATCDATRTGAA 23
 |||||:||||:||||:|
 DB 53 AATGCCCAATCAATATGGA 33

RESULT 10
 HSM005641/c
 ID HSM005641 standard; RNA; EST; 423 BP.
 AC AL041165;
 XX
 SV AL041165.1
 XX
 NI el399502
 XX
 DT 12-MAR-1999 (Rel. 59, Created)
 DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
 XX
 DE Homo sapiens mRNA; EST DKFP434G1816_r1 (from clone
 DE DKFP434G1816)
 XX
 EST: expressed sequence tag.
 XX
 KW Homo sapiens (human)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
 OC Primates; Catarrhini; Homidae; Homo.
 XX
 RN [1]
 RP 1-423
 RA Koehrer K., Beyer A., Mewes W., Gassenhuber J., Wiemann S.;
 RT
 RL Submitted (12-MAR-1999) to the EMBL/Genbank/DBJ databases.
 RL MIPS, Am Klopperspitz 18a D-82152 Martinsried, GERMANY
 XX
 CC Clone from S. Wiemann, sequenced by BMF within the cDNA
 CC sequencing consortium of the German Genome Project
 CC s1 sequence also available
 CC This clone is available at the RZPD in Berlin
 CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
 XX
 FH Key Location/Qualifiers
 FT source 1. 423
 FT /db_xref="taxon:9606"
 FT /organism="Homo sapiens"
 FT /clone="DKFP434G1816"
 FT /clone_lib="434 (synonym: htes3). Vector pSPORT1; host
 FT DH10B; sites NotI + SalI"
 FT /dev_stage="adult"
 FT /tissue_type="testis"
 XX
 SQ Sequence 423 BP; 143 A; 74 C; 77 G; 126 T; 3 others;

Query Match 72.2%; Score 16.6; DB 53; Length 423;
 Best Local Similarity 60.9%; Pred. No. 3.5e+02;
 Matches 14; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAATGCCCAATCDATRTGAA 23
 :|||||:||||:|
 DB 80 GAATGCCCAATCAATATGGA 58

RESULT 11
 C69479 373 bp mRNA EST 23-SEP-1997
 LOCUS C69479 Yui Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone YK364d8 5', mRNA sequence.
 ACCESSION C69479
 NID 92430835
 VERSION C69479.1 GI:2430835
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidae; Rhabditidae; Peloderinae; Caenorhabditis.
 1 (bases 1 to 373)
 Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
 Sano, M., Miyata, A. and Nishigaki, A.
 Expression map of the C.elegans genome
 Unpublished (1996)
 JOURNAL
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1294543.

FEATURES
 source
 1. 373
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(el489)"
 /db_xref="taxon:6239"
 /clone="YK364d8"
 /clone_lib="Yui Kohara unpublished cDNA"
 /note="dev_stage=varied, sex=Hermaphrodite male,
 tissue_type=whole animal."
 BASE COUNT 108 a 62 c 101 g 101 t 1 others
 ORIGIN

Query Match 71.3%; Score 16.4; DB 36; Length 373;
 Best Local Similarity 68.2%; Pred. No. 4.4e+02;
 Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 AATGCCCAATCDATRTGAA 23
 |||||:||||:|
 DB 173 AATGCCCAATCAATATGGA 194

RESULT 12
LOCUS T02063/C 319 bp mRNA EST 10-NOV-1992
DEFINITION *EST02784 Early embryo, Stragene (cat. #937007) Caenorhabditis elegans cDNA clone CEST83, mRNA sequence.
ACCESSION T02063
NID 9278544
VERSION T02063.1 GI:278544
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS McCombie, W.R., Kelley, J.M., Aubin, L., Goscochea, M., Fitzgerald, M.G., Wu, A., Adams, M.D., Dubnick, M., Kerlavage, A.R., Venter, J.C. and Fields, C.A.
COMMENT Caenorhabditis elegans cDNAs Unpublished (1993)
TITLE JOURNAL
COMMENT Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: arkerlav@tigr.org Seq primer: M13 Reverse.
FEATURES
source 1.319
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_1lb="Early embryo, Stragene (cat. #937007)"
BASE COUNT 86 a 56 c 66 g 109 t
ORIGIN

Query Match 70.4%; Score 16.2; DB 20; Length 319;
Best Local Similarity 61.9%; Pred. No. 5.5e+02;
Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 RAATGCCRAATCDATRTGR 21
:||||:||||:||||:
Db 251 AAATGCGAAGTCAATATGA 231

RESULT 13
LOCUS 221669 362 bp mRNA EST 20-APR-1993
DEFINITION HSDH3E12 Stragene cDNA library Human heart, cat#936208 Homo sapiens cDNA clone H23E12, mRNA sequence.
ACCESSION 221669
NID 938542
VERSION 221669.1 GI:38542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Aulfray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Pouliot, Y., Sebastiani-Kabakchis, C. and Tessier, A.
COMMENT IMAGE: molecular integration of the analysis of the human genome and its expression
JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genzentrum Muenchen Laboratorium fuer molekulare Biologie

Am Kioferspitze 18a, 8033 Martinsried, Germany
Email: obermayer@wms.biochem.mpg.de.
Location/Qualifiers
1.362
/organism="Homo sapiens"
/db_xref="GDB:D057168E"
/db_xref="taxon:9606"
/clone="H23E12"
/clone_1lb="Stragene cDNA library Human heart, cat#936208"
/note="Vector: pBluescript SK(+); Human heart cDNA library. Cloning vector pBluescript SK(+)"
BASE COUNT 88 a 103 c 108 g 63 t
ORIGIN

Query Match 70.4%; Score 16.2; DB 20; Length 362;
Best Local Similarity 61.9%; Pred. No. 5.5e+02;
Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 RAATGCCRAATCDATRTGR 21
:||||:||||:||||:
Db 111 AAGATGCCAAGCTATGTGG 131

RESULT 14
LOCUS D60457/C 283 bp mRNA EST 28-AUG-1995
DEFINITION HUM10E12A Clontech human fetal brain polyA+ mRNA (#6535) Homo sapiens cDNA clone GEN-110E12 3', mRNA sequence.
ACCESSION D60457
NID 9962096
VERSION D60457.1 GI:962096
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
AUTHORS Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takachi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.
COMMENT Fujiwara et al. (1995) Unpublished (1995)
On May 8, 1995 this sequence version replaced gi:801069.
TITLE JOURNAL
COMMENT Contact: Tsutomu Fujiwara Otsuka GEN Research Institute Otsuka Pharmaceutical Co., Ltd 463-10 Kagasuno Kawachu-cho, Tokushima, Tokushima, 771-01 Japan Tel: 0886-65-2888 Fax: 0886-37-1035 Insert Length: 1889 Std Error: 0.00 High quality sequence stop: 382.
FEATURES
source 1.283
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="22"
/clone="GEN-110E12"
/clone_1lb="Clontech human fetal brain polyA+ mRNA (#6535)"
BASE COUNT 92 a 50 c 47 g 88 t 6 others
ORIGIN

Query Match 70.4%; Score 16.2; DB 23; Length 283;
Best Local Similarity 60.9%; Pred. No. 5.6e+02;
Matches 14; Conservative 7; Mismatches 2; Indels 0; Gaps 0;

QY 1 RAATGCCRAATCDATRTGAA 23
:||||:||||:||||:

Db 277 AATGCGCAATCATGTAGAA 255

RESULT 15

LOCUS C61779

DEFINITION C61779 Yuji Kohara unpublished cDNA *Caenorhabditis elegans* cDNA

ACCESSION C61779

VERSION 92420484

KEYWORDS C61779.1 GI:2420484

SOURCE EST.

ORGANISM *Caenorhabditis elegans*.

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiida; Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderiinae; *Caenorhabditis*.

AUTHORS Kohara, Y., Mochizuki, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.

TITLE Expression map of the *C. elegans* genome

JOURNAL Unpublished (1996)

COMMENT On May 9, 1995 this sequence version replaced gi:802443.

CONTACT: Yuji Kohara

Gene Library Lab

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 0559-75-0771

Fax: 0559-75-6240

Email: ykohara@dbj.nig.ac.jp

Location/Qualifiers

1. 360

/organism="Caenorhabditis elegans"

/strain="CB1489 him-8(e1489)"

/db_xref="taxon:6239"

/map="12p"

/clone="YK2488"

/note="dev-stage=varied, sex=Hermaphrodite male, tissue-type=whole animal"

BASE COUNT 110 a 83 c 78 t 1 others

ORIGIN

Query Match 70.4%; Score 16.2; DB 36; Length 360;

Best Local Similarity 61.9%; Pred. No. 5.5e+02;

Matches 13; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 AATGCGCAATCATGTATG 21

Db 139 AATGCGCAATCATGTATG 159

Search completed: September 13, 1999, 15:45:57

Job time: 4045 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:20:17 ; Search time 329.34 Seconds

(without alignments)
415.234 Million cell updates/sec

Title: US-09-325-095-2729.SEO

VI

Perfect score: 1 GCGAGGATTCGCGCARGA.....RTGCCRAATCDATRTGAA 43

Scoring table:

IDENTITY_NUC 679419 seqs, 1590154680 residues

Searched:

Database :

GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_cm.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pl3.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sy.*
15: gb_sy.*
16: gb_un.*
17: gb_vt.*
18: em_fun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_hum3.*
23: em_hum4.*
24: em_hum5.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_st.*
31: em_sy.*
32: em_un.*
33: em_vt.*
34: gb_htg1.*
35: gb_htg2.*
36: gb_in1.*
37: gb_in2.*
38: em_ba1.*
39: em_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	25	58.1	7215	36	DMP13K68D X92892 D.melanogas

C	2	25	58.1	6831	37	DMP13K68D	7215 bp	RNA	INV	20-MAY-1997
C	3	22.6	52.6	155147	9	HS992D9				
C	4	22.4	52.1	2987	7	SOYPI3K5X				
C	5	22.2	51.6	124191	11	AC005153				
C	6	22.2	51.6	1999	17	OLVNEWYORK				
C	7	22.2	51.6	118431	35	AC007089				
C	8	22	51.2	1368	1	ENRPRVA				
C	9	22	51.2	11338	1	DNU02463				
C	10	22	51.2	128150	8	AC007661				
C	11	21.8	50.7	159941	34	AC006878				
C	12	21.8	50.7	69822	36	AC003121				
C	13	21.8	50.7	40293	36	CEIC45G9				
C	14	21.6	50.2	39090	7	CS39KBC1V				
C	15	21.6	50.2	35383	7	SCDNAIV				
C	16	21.6	50.2	2189	7	SCYD1197C				
C	17	21.6	50.2	3063	7	SCYD1029C				
C	18	21.6	50.2	1993	8	XSASF2A				
C	19	21.6	50.2	130381	11	HUAC003003				
C	20	21.6	50.2	2352	12	RATIGFBP3				
C	21	21.6	50.2	2058	12	RATIGFBP3A				
C	22	21.6	50.2	110000	34	CEY47H9_0				
C	23	21.6	50.2	110000	34	CEY53C10_2				
C	24	21.6	50.2	180821	35	AC007115				
C	25	21.6	50.2	38980	36	CER06C1				
C	26	21.6	50.2	39736	36	CER07E3				
C	27	21.4	49.8	79826	37	AC002443				
C	28	21.2	49.3	3248	2	AF058711				
C	29	21.2	49.3	45672	36	DMC22E5				
C	30	21.2	49.3	932	36	PFANGS				
C	31	21	48.8	2887	7	KITRLOC				
C	32	21	48.8	146505	8	F21M12				
C	33	21	48.8	68048	11	AC004764				
C	34	21	48.8	309026	34	AC006760				
C	35	21	48.8	42467	37	CELF26F12				
C	36	21	48.8	58753	37	CELF25C1A				
C	37	21	48.8	40907	42	AC007459				
C	38	21	48.8	26857	42	AF048728				
C	39	20.8	48.4	527	3	MONPRP1A				
C	40	20.8	48.4	10419	7	AB027513				
C	41	20.8	48.4	36849	7	SCX11RA				
C	42	20.8	48.4	3856	7	SCYLR130C				
C	43	20.8	48.4	95190	8	SOYPRXIN				
C	44	20.8	48.4	95190	8	AC007203				
C	45	20.8	48.4	88037	8	T2K10				

ALIGNMENTS

RESULT 1
DMP13K68D/c
LOCUS D.melanogaster mRNA for phosphoinositide 3-kinase, PI3K_68D.
ACCESSION X92892
NID 92113637
VERSION X92892.1 GI:2113837
KEYWORDS phosphoinositide 3-kinase; pi3k 68D gene.
SOURCE fruit fly.
ORANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 7215)
MacDougall, L.K., Dorn, J. and Waterfield, M.D.
A family of phosphoinositide 3-kinases in Drosophila identifies a
new mediator of signal transduction
Curr. Biol. 5 (12), 1404-1415 (1995)
JOURNAL MEDLINE
REFERENCE 2 (bases 1 to 7215)
MacDougall, L.K.
Direct Submission
TITLE JOURNAL
Submitted (09-NOV-1995) L.K. MacDougall, Ludwig Institute for
Cancer Research, 91 Riding House Street, LONDON W1P 8BT, UK

REMARK	Revised by submittor 04-DEC-95, revised by [3]
REFERENCE	3 (bases 1 to 7215)

TITLE	Direct Submission
JOURNAL	Submitted (20-MAY-1997) L.K. MacDougall, Ludwig Institute for Cancer Research, 91 Riding House Street, LONDON W1P 8BT, UK
COMMENT	On May 22, 1997 this sequence version replaced g1:1150926.

Source

```

/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
/dev_stage="embryo"
/dev_stage="4-8 hours"
/map="68D"
477..6107
/feature="P13K68D"
/allele="-"
/db_xref="Flybase:FBgn0015278"
477..6107

```

BASE COUNT	1973 a	1726 c	1738 g	1778 t
ORIGIN				

Query Match	58.1%	Score 25;	DB 36;	Length 7215;
Best Local Similarity	60.5%	Pred. No. 8;		
Matches 23;	Conservative 9;	Mismatches 6;	Indels 0;	Gaps 0

Oy 6 YGAYTTCGCNCARARAARTGCCRAARTCDATRTGRGA 43
:: : | | | | : | | | | | | | | | |
Db 4873 TGAGCATCGCCAGAAACTCTGCCAAGAICATGTGGAA 4836

RESULT	2				
LOCUS	DMU52192/c				
DEFINITION	DMU52192	6831 bp	DNA	INV	15-JUN-1996
	<i>Drosophila melanogaster</i>		phosphoinositide 3-kinase (cpl)	gene.	
ACCESSION	complete cds.				
NID	U52192				
VERSION	G1272419				
KEYWORDS	U52192.1	GI:1272419			
SOURCE	fruit fly strain-Oregon R.				
ORGANISM	<i>Drosophila melanogaster</i>				

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1 (bases 1 to 6831)
Molz, L., Chen, Y.W., Hirano, M. and Williams, L.T.
Cpk is a novel class of Drosophila Ptdins 3-kinase containing a C2
domain
J. Biol. Chem. 271 (23), 13892-13899 (1996)
96278830
2 (bases 1 to 6831)
Molz, L.M.
Direct Sublission
Submitted (22-MAR-1996) Lisa M. Molz, CVRI, University of
California at San Francisco, 505 Parnassus Ave. Box 0130, San
Francisco, CA 94143, USA
Location/Qualifiers
1. 6831

Source

gene

CDS

"organism": "Drosophila melanogaster",
 "strain": "Oregon R",
 "/db_xref="taxon:7227",
 "/chromosome="3",
 "/map="68D",
 148. .5778
 "/gene="P13K69D",
 "/db_xref="Flybase:Fbgn0015278",
 148. .5778
 "/gene="P13K69D",
 "/note="PtdIns 3-kinase",
 "/codon_start=1",
 "/product="phosphoinositide 3-kinase",
 "/protein_id="AAC67117.1",
 "/db_xref="Pdb:1j27d2420",
 "/db_xref="GI:1272420",
 "translation="MSMSADIDVCKOFDDLAKATLSTLSGHALDYSRNRKNGSGYGY
 OSYVAGADYDAAGRSLOLHOPRHSVHVLAIPENARSRSTPAGQNDILCEFA
 PTKSPKSSPESSGFLIEDLORPTNDSALVPMGVVASAIPDYGEPHDOORTLA
 OPTPGWAGVGVAVPTGADIPGSAAMPAPAPAOORPLNSELQRTSMPAKMAV
 PNAAVYKAVVPTGADIPGSAAMPAPAPAOORPLNSELQRTSMPAKMAV
 TAPTTSHHSQPSMHSSTFAEANGAIFPRKRVPTGVSSSSHTGNGHSPVPRR
 NDLDLNEHEDSRVLEAPDLPLNTGDLSDASTSYAAYDPDFLESGDAAQTQ
 SDMPENANMDKTAIVASVNGVLIMKRDGFLDPSSTSSQYAPVPESLKLANG
 TISIPPEPLPRNOCESNMAKAPVRSPOSSVILDTYSSIPANVLDKRTKTR
 YELSSDDQTDPELLEFYHNHAKVEKAPVPHDAPIVNGVAAEENYHMYDTSK
 VHPALNTGSLVLAASNGKRGVAGGAPVETGDDISVAOVAALASLEQGVKAT
 TDYAKPIGLLELVAIPRTSLQSLCEHNSFQLEKIDVHGGLGICSTANNOALART
 EDDLDLPEHLTPPEVQVITVYNNMLLELLEMEIDKESAAADVPRGRVYSCGV
 OAKVALICLGSIDYHEINACVADKLKICEDEKKTSTASAPELVSDIGDAOVV
 RSMDEQIKVAKSIRDAVDELANFVRVAIVKTPDYSTPIPLSCVSKTIVAGC
 ILSCHLRPLPNKMKEDFVLCVOIVYGRLLSPKPVNLGSDNIGGLFPLNFSAMV
 OHPTLITPRELRALFVLYGQAASBEQPNADONGERQVTTLELGCSTQLDPKVM
 CGPYLSTLAMPEDKMGKAPARGCHQPDCEVLSIEVPPGRIEPEKHEVAKP
 PHYFASDIAADQELMDLTLELTCYGTATSEREVEFMKRLIQSYGNALPHAAHS
 DYANLIDHALLHSMAPLSPGLSELTLLRPDARYKREKAVVLSMKPNDDVPEL
 LVQSLKHDTYKSSMAAPLISQSLSPSRHHNNHLYLHSLDDPHNSIGAMADPEL
 DESQVTVRYTRRNKMLRYALICEKMLQRTGTHRQCKLTTIAESYEAKESSK
 KRSLAAGDEVDLQPCPLVGELELQVSRNCSYFNSNPLPLKINFGVDAA
 SLPLAAGDQDQDDQLTLITNNKMLYELDLKVTFFGCAVPTGYSGLIEVLS
 AETLRKLTQVEGCKSEKRPIDFEKFGKQDMQNSPEKFNKSTPLSLCAGSYAV
 ISEKTRNNIMLKTSGLHFIDEFKFGKQDMQNSPEKFNKSTPLSLCAGSYAVIN


```

repeat_region      /rpt_family="MER2_type"
                    /15921..16228
repeat_region      /rpt_family="L1"
                    /16381..16626
repeat_region      /rpt_family="L2"
                    /18709..18781
repeat_region      /rpt_family="L1"
                    /18893..19442
repeat_region      /rpt_family="L1"
                    /19581..19797
repeat_region      /rpt_family="L1"
                    /19979..20033
repeat_region      /rpt_family="MER"
                    /20932..21230
repeat_region      /rpt_family="Alu"
                    /22451..22490
repeat_region      /rpt_family="MER"
                    /22642..22787
repeat_region      /rpt_family="L1"
                    /23587..23789
repeat_region      /rpt_family="MER"
                    /25713..25819
repeat_region      /rpt_family="MER1_type"
                    /25820..26073
misc_feature        /gene="GRB10"
                    /note="match to EST AA129408 (NTD:g1689294) zn83c02.r1"
                    /62129..26219
misc_feature        /gene="GRB10"
                    /note="match to EST AA129408 (NTD:g1689294) zn83c02.r1"
                    /26640..26937
repeat_region      /rpt_family="Alu"
                    /27447..27623
repeat_region      /rpt_family="MER53"
                    /28767..28844
repeat_region      /rpt_family="MER"
                    /28921..29309
repeat_region      /rpt_family="L1"
                    /34010..34297
repeat_region      /rpt_family="Alu"
                    /34861..34997
repeat_region      /rpt_family="L1"
                    /35458..35758
repeat_region      /rpt_family="Alu"
                    /35872..37519
repeat_region      /rpt_family="L1"
                    /38300..38850
repeat_region      /rpt_family="L1"
                    /39174..39265
repeat_region      /rpt_family="MER"
                    /39607..39664
repeat_region      /rpt_family="L1"
                    /41280..41542
repeat_region      /rpt_family="Alu"
                    /42628..42737
repeat_region      /rpt_family="MER"
                    /43326..43618
repeat_region      /rpt_family="MER2_type"
                    /44122..44319
repeat_region      /rpt_family="MER"
                    /45297..45597
repeat_region      /rpt_family="Alu"

```

```

Query Match      51.6%; Score 22.2; DB 11; Length 124191;
Best Local Similarity 51.2%; Pred. No. 1.4e+02;
Matches 22; Conservative 10; Mismatches 11; Indels 0; Gaps 0;

```

```

OY      1 GCGAGYATTCGNCARGARARATGCCRAARTCDATRTGAA 43
DB      29902 GGAGATGATTGTGGCCTGAAAAATTCAAAACGATATTAA 29944

```

```

RESULT 6
OLVNEWYORK

```

```

LOCUS      OLVNEWYORK 1999 bp DNA
DEFINITION Ovine lentivirus DNA sequence.
ACCESSION  L19198
VERSION    L19198.1 GI:309627
KEYWORDS   Ovine lentivirus (individual isolate Cornell, strain North American) DNA.
SOURCE     Viruses: Retroviridae; Lentivirus; Ovine/caprine lentivirus.
ORGANISM   1 (bases 1 to 1999)
REFERENCE  Campbell,B.B.J., Thompson,D.D.R., Williams,J.J.W., Campbell,G.S.G.
AUTHORS    and Avery,R.R.J.
TITLE      Characterization of a New York Ovine lentivirus isolate
JOURNAL    J. Gen. Virol. 74, 201-210 (1993)
MEDLINE    93155645
FEATURES   Location/Qualifiers
            1..1999
            /organism="Ovine lentivirus"
            /isolate="Cornell"
            /strain="North American"
            /db_xref="taxon:11663"
BASE COUNT 754 a 315 c 512 g 418 t
ORIGIN
Query Match      51.6%; Score 22.2; DB 17; Length 1999;
Best Local Similarity 55.0%; Pred. No. 86;
Matches 22; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

OY      4 GAYGAYTRCGNCARGARARATGCCRAARTCDATRTGAA 43
DB      1215 GAGAAAGACACTAAGACAAATGCAAAATGATATGAA 1254

RESULT 7
AC007089      118431 bp DNA HTG 16-MAR-1999
LOCUS         Homo sapiens clone Dj0616m23, WORKING DRAFT SEQUENCE, 16 unordered
DEFINITION    pieces.
ACCESSION     AC007089
NID           94417340
VERSION       AC007089.1 GI:4417340
KEYWORDS      HTG: HTGS_PHASE1.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE     1 (bases 1 to 118431)
AUTHORS      Waterston,R.H.
TITLE        The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE     2 (bases 1 to 118431)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (16-MAR-1999) Genome Sequencing Center, Washington
              University School of Medicine, 4444 Forest Park Parkway, St. Louis,
              MO 63108, USA
COMMENT       * NOTE: This is a 'working draft' sequence. It currently
              * consists of 16 contigs. The true order of the pieces
              * is not known and their order in this sequence record is
              * arbitrary. Gaps between the contigs are represented as
              * runs of N, but the exact sizes of the gaps are unknown.
              * This record will be updated with the finished sequence
              * as soon as it is available and the accession number will
              * be preserved.
              *
              * 1 1972: contig of 1972 bp in length
              * 1973 1990: gap of unknown length
              * 1991 3960: contig of 1970 bp in length
              * 3961 3978: gap of unknown length
              * 3979 7474: contig of 3496 bp in length
              * 7475 7492: gap of unknown length

```



```

FEATURES
Source
1. .1138
/organism="Dichelobacter nodosus"
/insertion_seq="IS1253"
/db_xref="taxon:870"
/note="A198; compl associated copy of IS1253"
186. .211
/note="nix1; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
199. .2043
/gene="ompB"
<199. .2043
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpB"
/protein_id="AAB12360.1"
/db_xref="PID:9407292"
/db_xref="GI:407292"
/transl_table="TLSKIDGLEDAYKKKOKELENONPADKVVDPQTQGVNGLDL
GKASOGTTOYRGESKPDVRNNEKEIASATSYGLENDONONPELNIYAKVND
GKVLIOYIGEXKAEIKDPKSLQONKFNPAKIDGYNLAKIDPAKIKGVKDYKIAI
ASYESLKDITNPDEKYLASKLQADLEASASESEYLYDPKHDSGRIGAVNYS
AQEFVLEGDAAQTQDAVLGTADGVSSVTEVAAPYDNGVATSTTGAASGLVGN
EKQETKRAIQTGTEESQSTRIRIPROYNDQAQISKFNYPAGAKAVIPSPTEFSLNL
FTGTGKTVKEDKHNDNIKAVGDLGADINKGTGVYHLGVKPKLEHVQYRVYTL
DPKLTADDEKSDAIKKRAAFSKNADKESLADNFEYRGTKATLTLOKALPKDRT
LVYQHALMGIONSFHGKSDNDVPSPADAPADNYSALGNFVANDYDKREYTGAV
YNWAKTEKETPPDKLYKEGGYFGYTVJAKAERTYKDKTNDADPRASFTGEKASAM
GGSNASTYDEKISSDDVWGVSFGAKLADPDKEDDNSFLGDKPKP"
complement(2115. .4031)
/gene="OmpE"
complement(2115. .>4031)
/gene="OmpE"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpE"
/protein_id="AAB12361.1"
/db_xref="PID:9407293"
/db_xref="GI:407293"
/transl_table="TLKQIAPBELAAKTEAEKERRKANLWVDPQTQGVVDGDLSSQ
ASYAGQYRGESKNEFDQNPENSAPTITGVLDONPKLTNIFPAKSDKDK
LVAIQYVIDPNKDVDLTDGSDSQONKHNVADRLALSHGTQPSDELEYANGQIQ
LOPSGAVLMLQAOILIDYNAVVDTOITIKKALASKELEQADKEARAVENKALIDKAL
ADLEKIKADKAKERTMLVLIDKYSKTIIPMGDYIPGSKILVNGDGAKALVVEGRT
GGVAGPLIANNPADIENRAVAYEKISGATTTLESNIOGTBENSYSSTRIFKGFSD
GTDVANSYTGATTVSEASLAERKAGATAGSTGLVGFADNKAATFDKQGFEEFYKSDS
NKVQANRHSILAAEALYLENVQYRGVSDIDAELAANKPEDEIKHIOSPYTRND
KAVQYFYRGNTNATTLKEMALPTDKILVYQGHMLMGIONSFHGKSNLSLAFGKG
KDNVTAIGNFRAEVDLNKDKVAGAVYVWQGETSELIPDLNVKFEENGVNTVYGA
ERTYKEGTADPRASFEGKEANMGGSFNSMLANDGYIHEDGVWGVSFGAARTYEPYK
ESFYDNKWLISHG"
4019. .4044
/note="nix2; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
4332. .6460
/gene="ompD"
4332. .6460
/gene="ompD"
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpD"
/protein_id="AAB12362.1"
/db_xref="PID:9407294"
/db_xref="GI:407294"
/transl_table="MKLSHLAIIISAITLAAGCGNDKPMPODNGALVBOBNA
IKKATAYLEDKDGKAPALTKDEANTLSQIAPELAAKYKKEKERRKANPDADVYDPT

```

```

sig_peptide
/misc_recomb
4316. .4542
/gene="ompD"
/evidence=experimental
/product="110protein signal peptide"
4516. .4542
/gene="ompD"
/note="nix3; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
complement(6534. .8357)
/gene="ompA"
complement(6534. .>8357)
/codon_start=1
/transl_table=11
/evidence=experimental
/product="OmpA"
/protein_id="AAB12363.1"
/db_xref="PID:9407295"
/db_xref="GI:407295"
/transl_table="TUSQIDENLFENYKKKOKELENOASDDEVDPTKGVNGLDK
QAOQGTTOYRGESSEFDEANNKREVSATALLGVAKQONPMLNVIARVADGT
VSIQYIOMGNDSDAADFTNYSLOQONDNVNVKISDLKTNKPLDEAKLEAKROVDLI
IKLYEKIDSEITTEKEELEKFEALKEELNADIDSPKGYIDDEKAKLSEGCAT
GSDDIIFGDDVPNNNSVLGTADGSGAANPVPIYNSGKRGADATGAGLGVNMFK
VGTGQALIGTEKSSSTRVGRQYNTQSRAPAKNSYADAAVIALSSEKIDANAI
GTTQSELANITIDKRNKINHLAVKPKLDYVOIGRVGNLDPIKVSLETDDEKAK
LEKGETITKRAAFADPKGERADNFEYRGIDATTIEQMAALPTDKILNYQGHMLY
GIDNSFHGAKSQSFSTAGTAGALHAIGNFVSAEVDLSKGVGHYNNWADGSKDRV
IDSLVFRGVNHTVIGTAERTYKAGDADPRASFEGEAGMGGSFNSVTTAEKYS
DVMWGVSFGAKRIVERPKSDDNGFGLDPLDTTKPKPKP"
8345. .8370
/note="nix4; cross-over site"
/organism="Dichelobacter nodosus"
/evidence=experimental
8664. .10361
/note="IS1253"
complement(8684. .8722)
/note="terminal"
complement(8735. .9088)
/gene="ORF117"
complement(8735. .9088)
/gene="ORF117"
/codon_start=1
/transl_table=11
/product="ORF117"
/protein_id="AAB12364.1"
/db_xref="PID:91488248"
/db_xref="GI:1488248"
/transl_table="MLVTYKRRKCFISAMLDLKLKYLDLCKWEVEYLEFNGADHY
HLLDMYPIIMPISKFTINLKTYSRLMRKEFEHLAEFWKPVLMTRAYCLLTGTGAT
IDTIRQYIEKQERP"
9089. .9099
/rpt_type=direct
9100. .9110
/rpt_type=direct
9110. .10237
/gene="ORF375"
9110. .10237
/gene="ORF375"
/codon_start=1
/transl_table=11

```



```

gene
985..4254
/gstraln="Bristol N2"
/gene="C4569.6"
join(985..1062,1269..1451,1635..2019,2534..2661,
2710..2827,2932..3120,3173..3417,3656..3781,3865..4254)
/gene="C4569.6"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62549.1"
/db_xref="PID:9687871"
/db_xref="GI:687871"
/translation="MTALLLICSTITPSSDGVVIGGGVGVQVQIRANPLSFSSCD
HLLVTFEFSNNHNGTGHVDCCTTITFCNKLCCCLMKGPIIDPGNONNPSPTNG
GIGGVYFPGGGNGGVQDEEGYICNSNDCKSLGCLTASVGVKICLSTSGGGG
FPGNGGCQTSSNCQGVYVNGNQIQTGGVSPARQMWAPSSNSITVDEN
EDVFSKMAPEGRKINGECRADCDDELCTMYFGEMCRSPKPLIPRCSDAEP
STEYLCVFSTAMADRYCYKGVYVIGVYIPKHKISMELKSTTSSPTTTHLAK
PESNGLFAESEAFIEQPGSLTSLALOKRADMSONGPTPMYVTKMSDIPSHVLAG
HGEAVRITKVVYKKEKEMENGEREYKRIPIKIGVGVVDEPIDPATCKGESCS
GVYRVDNRVYICRTDMFKHRCCLHSDCISGRCPTGRDIACTEDTISATISIQ
CFYDECSGECCTLVDEKERKFCNPSPTSPRANQIOTNSOCPEQVCAEKNNAE
LKITISMVIFESNFTONSESYGCFSONSKMLSVPTYEESVSVI"
complement(10554..11408)
/gene="C4569.7"
complement(join(10554..10673,11154..11408))
/gene="C4569.7"
/note="similar to tight junction protein (ZO-1)
(SP:Z01_HUMAN, Q07157); coded for by C. elegans cDNA
CEES55F"
/codon_start=1
/protein_id="AAA62550.1"
/db_xref="PID:9687872"
/db_xref="GI:687872"
/translation="MTAYGHMGEALIECLISAVELHKOEVIDAHQVYIRVFKLGGG
IDDPPTKAPFKYPDGSGVITVYVSGSPADVAGLRKHDKILQVNGADFTMTDRAYKE
IKOSVYHLVLRADLPVSLH"
12390..14157
/gene="C4569.5"
join(12390..12512,12560..12679,12728..12827,12875..12959,
13551..13688,13773..14157)
/gene="C4569.5"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62548.1"
/db_xref="PID:9687870"
/db_xref="GI:687870"
/translation="MTOEPSDTSYQREDEMLLHEPGARVPYVNVYDQNNIVAS
GGLYFNGKTLRTLHEDKKFKMAPPOVSYNDRIITYSAEPSDRTILRVSLD
GVVHESFEGLLNGMTITDGDWYIVKQAEGBHNNMTHTLDTPIGMQDVVASEK
GEAFSRICADDKTIVASINDPMMNESQRIYFDLENTETGSTRAGKRPGEITF
PFIKRYGDFILNDKSGRSEFKNGEVALIRAKIDAVLGEGFVKGGEALMALITGM
VADPNEQLICDWMLETIKLDGSSKRAKREKKOEDDOAKN"
complement(14436..18535)
/gene="C4569.8"
complement(join(14436..14510,14558..14863,14917..15332,
15809..16070,16985..17076,17435..17503,17552..17604,
17689..17755,18478..18535))
/gene="C4569.8"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62551.1"
/db_xref="PID:9687873"
/db_xref="GI:687873"
/translation="MRLVCGFFSFYFSSRKGRSYFRWVSTETSNTRSKCHLK
SSSITLISLMQIDRLRSTEHVEIYPOETCVHNSPDVLTLECATMSTYKMP
NSSYISKSATKKTGSKKVTYRSKSKTKTITGNSKTVTTTISNKPVDKXPS
KDOORQTKRQDPEKEMWKECDVTRCKVIOEKFIAEQLRIBEDONETYSR
DIDEADMERICKKFAKEKINPTTAAEPIDDKITLLDRYTKKPYPLAKTNNEGCLIFD
EBSYFSKNEKKAKSESNSLEEDTYGVQPKIDVQRPQNVFSRPGVPPAVVLLP
TEEDLVADPSISVGETHELMYHLKQVPIQTKSKLILNALQPSILEDRODHFHP
ELVEFNITRSLVHAQEMEGKNEKSDEGKSKSLSFVIEKOESEFYSRAPITETAR
MARRNRSGSSSKK"
complement(18908..20020)

```

```

CDS
/gene="C4569.9"
complement(join(18908..19103,19235..19661,19710..19944,
19994..20020))
/gene="C4569.9"
/note="weak similarity to C. tentans Balbiani ring 2.1
(PIR:A45294)"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62552.1"
/db_xref="PID:9687874"
/db_xref="GI:687874"
/translation="MTNATPDDKKIVSPKPTKTTSDKSKTKPRRSSTSKRKSCK
LFCCAKRKTTRKSKSKRTKRSAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK
QSGVDAQATVTPRPPTPIPTGVKPEAPRSEPLQPRSVSTTPRSATGTTP
EOMVAPATLTPPVSABSKHLPODPGDASSRVROYAEYSKEDQDDQDKLRKS
VAYPRHKFFPMQYVDECRVRRVYEDSVPLMESNMKMLRMAASNRIACQSDKAR
CDMKDLMEHIIIDNGN"
complement(21146..26146)
/gene="C4569.10"
complement(join(21146..21383,21527..21953,22344..22487,
22554..22840,23258..23618,24037..24180,24454..24740,
24830..25419,25464..25591,25745..25906,25956..26048,
26092..26146))
/gene="C4569.10"
/note="coded for by C. elegans cDNA yk286c.3; coded for by
C. elegans cDNA yk26e8.3; coded for by C. elegans cDNA
yk286c.5; coded for by C. elegans cDNA yk26e8.5"
/codon_start=1
/protein_id="AAA62553.1"
/db_xref="PID:9687875"
/db_xref="GI:687875"
/translation="WMLNLTVPATSVLLAEGTFDNCNGLLTDKILIKTSPGNVS
VMDINEVTFGEBSGKLSLIPROFSIKLEKRVNTPLVNVLPIRBDYRHNVTYEX
IVSKEMWHSIQFDLOSILKLTVDKSGSKDLILYDLRAEGQOTSKILLAPNRDP
CORPAMQDITIRSOVQVONAMRNSHVSQSGRTAFLNLTGDEPNYVTKNEIRL
LNYKILNANYAYLVVEETILKPGKAEVVOVORISKTREREYGVPTIVPEIDF
KXSSFSNENINPNQISIFKTPLYDPRKNALSPDGEVYININRSPITSELYPV
KTENKHFQPMASOEGIFRYIRHDSNTLITIPFDIRNELSEENRLAYCYTMD
LYFVNTITDGVQNSNNITVOMKKSSTNSIESIMLIPARFQYVDTVMIIYEEHDGL
LRFVGENMTIEVKTYKTYLRETLVDKXVILNFGEPYNTSVDIISNNNSKV
ETASVOINGKSGSYIRIGNRYVDYFIFNPSSEWNVDESAGVGMFLNDPFR
KKSIGKPDENNNDIORSPDYDGLVIVSNKTKNDGPAITVRSLETGILYEHQY
ENPMYFSIRPQIRNELPVEHLSIERDPTVYHVITKKNHSIGTSEVNSFNWEX
QENACTIAEYGVHNTYINGETILUEBSKSLFDPKNAVITVILLSSKCSDLVN
REFANKFSYQVLEWCPDSKGYEDITYVKSNNQVKNVYQIGSRYEKNQTSPT
YIPVTSYPTTPNPSPALPMLIDFLNERNENVEKXSTELAMICSEIENGCTTI
TRVLPVNTYITIAKLTGKTGIAFINKRSQTAAYRLTEKSTMGFEVNSVSNHEG
NEWYTSFGEORKNKRESILATLENNSKRTKEVNVQVAYKRSFLDAKFKLSS"
26971..28084
/gene="C4569.4"
join(26971..26997,27047..27281,27337..27577,27889..28084)
/gene="C4569.4"
/note="weakly similar to C. tentans Balbiani ring 2.1
(PIR:A45294)"
/codon_start=1
/evidence-not_experimental
/protein_id="AAA62556.1"
/db_xref="PID:9687878"
/db_xref="GI:687878"
/translation="MTNATPDDKKIVSPKPTKTTSDKSKTKPRRSSTSKRKSCK
LFCCAKRKTTRKSKSKRTKRSAPKAPKAPKAPKAPKAPKAPKAPKAPKAPK
GIYDADAKATVPRPPTPIPTGVKPEAPRSEPLQPRSVSTTPRSATGTTPD
MTAPATLPPPSABSKHLPODPGDASSRVROYAEYSKEDQDDQDKLRKSV
YPSHKFTWQVQKDECRVRRVYEDSVPLMESNMKMLRMAASNRIACQSDKATCD
MKMLNTEETIIDGNF"
complement(28258..30680)
/gene="C4569.11"

```

Query Match 50.7%; Score 21.8; DB 36; Length 40293;
Best Local Similarity 53.5%; Pred. No. 1.8e+02;
Matches 23; Conservative 8; Mismatches 12; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCAGARAPARTGCCRAARTCATYRTGRAA 43

DB 7566 GGGGACAAAAGTGTACGAAAAGTGTCAAAATTTCGTGTGAAA 7608

RESULT 14
CS39KBCIV/c
LOCUS CS39KBCIV 39090 bp DNA PLN 07-MAR-1997
DEFINITION S.cerevisiae 39kb DNA segment of chromosome IV.
ACCESSION X99000
X99000
X99000.1 GI:1429338
NID 91429338
VERSION X99000.1 GI:1429338
KEYWORDS anti-silencing protein; Asf2 gene; GABA transporter protein; gdh gene; Gly-tRNA; HEN3 gene; high-mobility-group-like protein; hydroxymethylbilanesynthase; methylated-DNA-protein-cysteine demethyltransferase; MGT1 gene; NAD-dependent glutamate dehydrogenase; NHP2 gene; secretory component; SHM1 gene; SHR3 gene; sugar transport protein; UGA4 gene.
baker's yeast.

SOURCE
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 39090)
AUTHORS Bahr, A., Moller-Rieker, S., Hankeln, T., Kraemer, C., Protl, U. and Schmidt, E. R.
TITLE The nucleotide sequence of a 39 kb segment of yeast chromosome IV: 12 new open reading frames, nine known genes and one genes for Gly-tRNA
JOURNAL Yeast 13 (2), 163-169 (1997)
MEDLINE 97197981
REFERENCE 2 (bases 1 to 39090)
AUTHORS Schmidt, E. R.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1996) E. R. Schmidt, Johannes Gutenberg-Universitaet, Institut fuer Molekulargenetik, J.J. Becherweg 32, D-55099, Mainz, FRG

FEATURES
SOURCE Location/Qualifiers
1..39090
/organism="Saccharomyces cerevisiae"
/strain="FY1679"
/db_xref="taxon:4932"
/chromosome="IV"
/clone="cosmid 1C10"
/map="left arm"
complement(636..1259)
/note="ORF D0884"
/codon_start=1
/protein_id="CA67473.1"
/db_xref="PID:e253394"
/db_xref="PID:91429339"
/db_xref="GI:1429339"
/db_xref="SWISS-PROT:Q12328"
/translation="MYTGFLEQISPAOKKPYNELPPEOGERGAEAMINMPTSCPG KSVYGVTFALGVGLGFMSMAVDTPLHTPTNTAATADAGNIGVGSRTYQOI SDLPROOMKLOFTDMGRKYSASAKNPGYICMIYAGCVIYESLARNDIYNGVTAG FTGAGLAKAKAPQALMGAGFAAFSAIDLYKMSGDGRPNODEKE"
complement(1650..3017)
/note="ORF D0888"
/codon_start=1
/protein_id="CA67474.1"
/db_xref="PID:e253333"
/db_xref="PID:91429340"
/db_xref="GI:1429340"
/db_xref="SPTRMBL:Q12468"
/translation="MNNDRKSIKTAGGSMLSNKTVKELNQLKERYTVEDELTESTI ALSWRFPQEPDEPHALSSSLKTKLKQOOSTDIPSTHVLISLSECKITHYAVR GGNIEWGIILGFTLKNIIVVDCNLPVNGTEPPYNAOLSEYMYOYIDEMAHND GGDGRDYGAKLNTVGRFHSHPGIDCMSTNIDIQODLNORPODYVAIVDDPLKSLK DKLWGAFRITESKSDNSATSYIELTIIIFDSLNLALPEFKLNLMLCVLEDESEQ ISLNLIDSMKQYSLYMDKSNVTRIKIATSERVSNENKNKIDONSTSCFOCIANT ORGDSLETSEFGSMFSGDNTSDVDEMDRNLTEFDSTSLCINGEPIHVNREVERSR STDNENHKKRRNMSQERCHDEGNDMLQRVLETDYAKAKRILASKIKOYERLRFYK DTFPL"
complement(3293..6571)

gene

CDS
/gene="GDH"
complement(3293..6571)
/gene="GDH"
/note="ORF D0892"
/codon_start=1
/product="NAD-dependent glutamate dehydrogenase"
/protein_id="CA67475.1"
/db_xref="PID:e253395"
/db_xref="PID:91429341"
/db_xref="GI:1429341"
/db_xref="SWISS-PROT:P33327"
/translation="MLDNKRGKLNLSLNPDIASLSISMSDHYVDFPKDQREBE VIDLDDQGFIPDILIQEDVWFNLSGLIDLFPSRSPQILSTIHSLSAKLDLFFA KSKNGQPLRFSIKNITINDNHAIFMESSTGVISDSQCNKRFASDVGNTLVEH GKDPIKRNRIEMDSCCPYELSDIDFLDNKQKCNPLVFAPAPESLKEPLFVYS VYPDDPAGVVISODLKGDIESISQKTYMKVSNENKLYGLLILVREBPVYK TTRVENKDEIRILIAVAKRPTTKRYEYVATKLPKRFPSFNPVDDIITIF SVYLNENQLEDVLLHVEALAKOYERASLLIAPNNSPHEVYQRPQSFKEIYAH IGAIFNPNVNRGSDYQNLSTQITIRNDPTLLEIVENLKRKLNETLTOOTIINM SKATYIISLKLKRNQAIHYHNSTKDMEKLSFQRLKEVEFKNDQDEEALNFIEN DSPDLLILKTLNIFNKSILKTNFIRTRVAISFRLDSLVWTKFEYETPGIIFVVG NTFKGFHREFRDIARGIRIYCSRNODIYDLSNVDENYQIASTOORKKRDIPEG SKGYILLNPGIYVEDQTFVAFSOYVADMIIDLINDPLKENYVNLIPREILFEGDDG TAGVVDATNARARVNRKSPKSFITGKSPSLGGIPHDEYGTSGVRAVYKITYETLN LTNSTYIKFOTGGPDGDSVIELISSPNECYIALILGSGVLDCKDQKDELRLH ERKMSIDFISLKSNNGEFVSDAMDMLPNGTIVANGTFRRNFHQIIFKFDVHVI FVPCGRPNSTILNNLHYFDEKTKCKIPYIVGAMLIITOPAKNALEBGCILFED ASANKGVTSSEMEVLASLALNDNDPEHFKFIDVSGERSALYKSYVAVOSRIQKME LEFQIWNLNQINGTHTSETSNOLFTIINKLNDLVASQELMLNDLRLNRLDKLII PKLIIDVAGPOSVLENIPESILKVLISLSYSTVYONGIDVINIGKLEFIIGLKREA EASA"
complement(7099..9198)
/note="ORF D1014"
/codon_start=1
/protein_id="CA67476.1"
/db_xref="PID:e253334"
/db_xref="PID:91429342"
/db_xref="GI:1429342"
/db_xref="SPTRMBL:Q12310"
/translation="MSLSRLRYNRNNKTTASLTAEHAYSDNNAYSVSLDPTSGV MAARKGEALNKSYDVPSLPVADSVPRDTFTSSSDDEDTOKLTTSMKEDILF MPENISKIDSNISSPAYSRKAEAFIYPKGRKLRNSDFTSPENSLSRSTYKILF FTSKNSSTSKNLANDISDNKHNNSNTIGHIPIYTAATACDELACTSIEEYVY YEEERMFTRYVSLSDSVSLSTPLDQTSYEAQVPTRIHLEDESTAHIRKHSYTS LLSIKRLFTIRYFSNNSNSCSDHOEYVADCAISSIKETSPVSTGSPSLIENE DSDRDOIIOALYSINEASTDLVSRKYDLVVLGEGSGKYLQORVLDNKFAIKEX BSKRRRSEERKYIKNITSEVCIASLTANPNPCLELITIEYERKIFOLLECEYDLSL VMSERMAHEICILFKQILINCYKTLHDIGLSHRDLKDKNCYTRRGILKLDIFPAA SV FHYPLSSOMIEANGIVGSDPYLSPVEYFYENYDRALDVMVSGIIFPCMITRRPMTK PKRVDOVFKAFCSGRGSFENDLVTRATDSDSNYDNDGVEGVIDMGPMITLARLE ETHIRMRILEVSEFFRITINGIIDLQDWIKIEFTCOVGAASPNEASLRILINKNHHI TTIORRAHIGLGHOR"
complement(9942..10619)
/note="ORF D1018"
/codon_start=1
/protein_id="CA67478.1"
/db_xref="PID:e275392"
/db_xref="PID:91628361"
/db_xref="GI:1628361"
/db_xref="SPTRMBL:Q07623"
/translation="WGSBEDKLLTKKOLAKAQFRKSEKEDQERKVEKQAPCKRRP SAAQNDSEEPYKRRKTRRGKRGKNGKKNRPIYVGSJPRDITAVELONHKNS PQQRLRADKGIAPLEPADPDRFGIORMDIALLOGLTLKEKINVELVGGGNS QERLEKLNKNIKIDLEERKEKLTLMINDGNQKIAKTATATAORTSGDNKRVVPGIHP DRATLKL"
10237..10539
/note="ORF D1019"
/codon_start=1
/protein_id="CA67477.1"
/db_xref="PID:e253335"
/db_xref="PID:91429343"
/db_xref="GI:1429343"
/db_xref="SPTRMBL:Q00590"

CDS

gene
complement (2986, .3888)
/gene="D1214"
complement (2986, .3888)
/gene="D1214"
/codon_start=1
/product="suppressor of abf2"
/protein_id="CAA58249.1"
/db_xref="PID:g1004297"
/db_xref="GI:1004297"
/db_xref="SWISS-PROT:P38988"
/translation="MPRTDKKQSLARLGLSAGSLMEIAVHPVDITISKRLMSNHTK
ITSGOELNRYIFRDHSEPLGKRLFTLPGLGYAAYKVLQRYVKGQGFANFLK
HYKDFNLFGKTKGKAMRGAAGSLIGIGIYLLPLDVLKIKQTPSEFGKGFYK
ILRDEGLFNLRYGKWTAAARNAPGSFALPGSNAPAKYIILGADYSQATSONFISI
VGACSLIYASAPLDVITKRIQNRNFDNPSGLRIYKNTLNKNGVTAFFKLTPLRLTT
GPKLVSFALAGSLIPFDNLK"

gene
complement (4254, .5831)
/gene="D1219"
/codon_start=1
/db_xref="SGD:S0002356"
complement (4254, .5831)
/gene="D1219"
/codon_start=1
/product="anti-silencing protein"
/protein_id="CAA58250.1"
/db_xref="PID:g1004298"
/db_xref="GI:1004298"
/db_xref="SWISS-PROT:P32448"
/translation="MPKRGVLAITRSVIDSGDESSSFNSDKREYAAYTKGLSSR
VOKKSIRKQMSKVKELQSLVNHRENEALVSSAKLSCGELIGYEIKMASLHGKMS
ILDENNLKTEKHSASAKRIELVLPSSKEBNYDETLVNLKKEICAKIDYKKNQ
NMYNTKDEIHTEHEKYEGLEISIDKVPDASSKELAVRRETNVNRRETEKYNV
LKMOLOATKSELEKQAKADYKLCIPELVKANTLHLSYKSTANRRETEALQ
AELSQOSETGOLETQNOVOIPTNTVLPDPEENNPEDLFAIOBELDOLRLHKM
ADERSRTHLLEKRNKIKRLQSYQSLQRLPPAKRHLGIQRLGSEKSCPLP
AVASTYAPLLSHQSNHCEIDNTPQRLLAAPGQCYSEKSTYLNLDYSRKSYSLR
LOPPHIANLKSLETKTLPVPPTDSPLQPSKDKSQETAKKDDRPALVNEPVTLDTSRP
PVQASLADSKHSCGLHK"

gene
6078, .6407
/gene="D1224"
/note="putative ORF"
/codon_start=1
/protein_id="CAA58251.1"
/db_xref="PID:g1004299"
/db_xref="GI:1004299"
/db_xref="SPREMBL:O12187"
/translation="MPSESRRLSKKGATVRLVENVYICFTVLCFCLIPHSID
WRSGLSGYFIDFFFFHLSIPRWTFYPSVKNHHTSIRPRKSEKNKQVSDPFLX
SCSRVAF"

gene
6545, .10366
/gene="D1229"
/db_xref="SGD:S0002354"
6545, .10366
/gene="D1229"
/note="putative ORF"
/codon_start=1
/db_xref="SGD:S0002354"
/protein_id="CAA58252.1"
/db_xref="PID:g1004300"
/db_xref="GI:1004300"
/db_xref="SWISS-PROT:P38968"
/translation="WKLAEFSRTAFAMSHDKIPLVSGTGVTDANFSTDSLEL
WSLIAADSEKPIASIQVSKFNLDWGNHKKIIGALDNGSLSELYSTNEANNAINSMA
REFSNHSSSVTKFNNAKQDNYLAGNGGEIIFIDANKCTSPSNYTPPLPGOSMSY
DEVISLAWNOSIAHFASAGSNFASINDLAKREYIHLSTSPNSGIIKOOLSYVENH
PKNSTRYATNGSDNPSIILMDRANNTPLQTIHQHOKGILSIDMCHODEHLLLS
GRDNTVLNMPESAELQSLQPPARGNMCFTKFAPEADLFCASFDNKIEVOTIQLNLT

gene
11115, .13571
/gene="D1234"
/db_xref="SGD:S0002353"
11115, .13571
/gene="D1234"
/codon_start=1
/db_xref="SGD:S0002353"
/product="High affinity glucose transport protein"
/protein_id="CAA58253.1"
/db_xref="PID:g1004301"
/db_xref="GI:1004301"
/db_xref="SWISS-PROT:P10870"
/translation="MNSVPTDDISTIDNLSLSEPPROKQSMMSICGVFVAGFL
FGIDGLINSITSMNTVKSHPVADNHSFTNQQSILVSLSTCTFGALTAIPISY
GRNPITLITSTIFLIFSGNSLQVAGGITLLIVRVSIGIGIAISAVPLPQAEATHK
SLNGALISTYOMITWGLVSSVSGTHARNASSRPIGIGIYWSFLAIGMFL
PESPRRYVLDKIDDEAKSLSEFGRVPHSGLELVEIKATDYDEASGSEFTDC
FISKSPKOTLMEFTGIALQAQOQSGINEIYGVNFKNGTVGNSYVSPITVAV
NYPANVLPETFPVPPRRKRYLVGYLMTANIVAVIGSLKTVAAKRWIAFICLF
IAAFSATGSGVAVINAEIPLVRSKTRICAAANLVNFTALITPIYVDGSHS
SLIAKLPFLMGSINAGVIVLYLVETKGLTILVREIDELYIKSTGVASPKFVKDIRE
RALKFOYDPLEDKNTEFVARNNDETPENDRMTISGDIHSPKQEVHSIPE
RVDPIDPSTELLESPPKSSGWTVPVPSLDQVPJPOTTEPAEIRTKYVDLNGGLNTY
NRGPPSLSDSSDYDEIDEGPSOGDSNRSTMDINDYMARLHSTSTASNTDG
FSGNOSTLRVHTASHSIDTPEEDPSIMDGLNGIALNAYNNGPSIILMSDEANNGE
TSDNLTADQLACKERMAQFQSYIDKRGGLPEPQSNILSTISLVMADTNHNEI
LHSEENAINQPVNENDLK"

gene
14009, .15136
/gene="D1239"
14009, .15136
/gene="D1239"
/note="putative ORF"
/codon_start=1
/protein_id="CAA58254.1"
/db_xref="PID:g1004302"
/db_xref="GI:1004302"

Query Match 50.28; Score 21.6; DB 7; Length 35383;
Best Local Similarity 51.4%; Pred. No. 2.1e+02;
Matches 19; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
OY 6 YGAYYTRGNCARGAARAGCCGAAATTCDAATRTGGA 42
DB 5218 TGAATTAACGTGAGAAAATTCGAAATCAAGTGAA 5182

Search completed: September 13, 1999, 16:20:31
Job time: 4543 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:22:05 ; Search time 133.08 Seconds
(without alignments)

80,841 Million cell updates/sec

Title: US-09-325-095-2729.SEQ

Perfect score: 43

Sequence: 1 GGNAGYATTCGNCANGA.....RTGCCRAATCDATRTGAA 43

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database: N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	25	58.1	6831	1	Phosphatidyl inositol
2	21.2	49.3	4488	1	Staphylococcus aureus
3	20.8	48.4	446	1	Expressed Sequence
4	20.8	48.4	446	1	Human brain Express
5	20.8	48.4	1269	1	Metal-regulated tr
6	20.8	48.4	2027	1	Homo sapiens SDF-5
7	20.8	48.4	1960	1	Human secreted pro
8	20.8	48.4	1869	1	Human secreted pro
9	20.8	48.4	1984	1	Human ATG-1622 CDN
10	20.8	48.4	1984	1	Human secreted pro
11	20.4	47.4	2289	1	Human secreted pro
12	20.4	47.4	2187	1	Tumour cytotoxic f
13	20.4	47.4	2395	1	Hepatic parenchyma
14	20.4	47.4	1661	1	Human hepatocyte g
15	20.4	47.4	2187	1	Human hepatocyte g
16	20.4	47.4	2172	1	Human hepatocyte g
17	20.4	47.4	2393	1	Plasminogen-like g
18	20.4	47.4	2214	1	Human hepatocyte g
19	20.4	47.4	2199	1	Human leukocyte-de
20	20.4	47.4	2184	1	Human leukocyte-de
21	20.4	47.4	2289	1	Human hepatocyte g
22	20.4	47.4	2394	1	TCF II-encoding se
23	20.4	47.4	1661	1	Human HGF gene par
24	20.4	47.4	2187	1	Human HGF gene par
25	20.4	47.4	2173	1	Complete human HGF
26	20.4	47.4	2184	1	Encodes haematopo
27	20.4	47.4	2187	1	Hepatocyte growth
28	20.4	47.4	2187	1	Vasculin endotheli
29	20.4	47.4	2187	1	Competitive inhdbi
30	20.4	47.4	2184	1	Competitive inhdbi
31	20.4	47.4	2187	1	Hepatocyte growth
32	20.4	47.4	2214	1	Human recombinant
33	20.4	47.4	2199	1	Human leukocyte-de
34	20.4	47.4	2172	1	Human leukocyte-de
35	20.4	47.4	2187	1	Nucleotide sequenc
36	20.2	47.0	5340	1	Nucleic acid encod
37	20.2	47.0	6869	1	S. cerevisiae scsu
38	20.2	47.0	4434	1	Saccharomyces cere
39	19.8	46.0	9623	1	Fumonis-in-resistan
40	19.8	46.0	17250	1	Staphylococcus aure
41	19.6	45.6	328	1	Enterococcus faeca
42	19.6	45.6	204	1	Encodes exon IX of
43	19.6	45.6	2042	1	Human gene signatu
					New promoter used

ALIGNMENTS

44 19.6 45.6 8033 1 X13058
c 45 19.4 45.1 447 1 T27954

Enterococcus faeca
Hepatitis C virus

RESULT 1
ID T80200/c
AC T80200;
DE 22-JUN-1998 (first entry)
KW Phosphatidyl inositol 3-kinase cdk CDNA.
KW antagonist; inflammatory joint disease; cell proliferation; cancer;
KW psoriasis; restenosis; atherosclerosis; therapy; diagnosis; cpk; ss.
OS Drosophila melanogaster.
FH Key
FT location/Qualifiers
FT CDS
FT 148..5778
FT /*tag= a
PN WO9731650-A1.
PD 04-SEP-1997.
PF 12-FEB-1997; U02193.
PR 29-FEB-1996; US-609049.
PA (RECC) UNIV CALIFORNIA.
PI Chen Y, Moiz L, Williams LT;
PI P-PSDB; W38757.
PI P-PSDB; W38757.
DR New isolated phosphatidyl inositol-3 kinase polypeptide - used to
PT develop products for diagnosis and therapy, particularly for
PT proliferative disorders, e.g. inflammatory joint diseases, or cancer
PS Example 1; Fig 9; 77bp; English.
CC This cDNA sequence codes for cpk (see W38757), a Drosophila
CC polypeptide that belongs to a novel class of phosphatidyl inositol
CC 3-kinases that contain a C2 domain, are capable of phosphorylating
CC a D3 hydroxyl of an inositol ring in phosphatidyl inositol and
CC phosphatidyl inositol 4-phosphate, but not in phosphatidyl inositol
CC 4,5-bisphosphate, and which are involved in cell signalling cascades
CC that control e.g. cell cycle progression and intracellular protein
CC sorting. Short fragments of cpk cDNA were obtained from a
CC Drosophila cDNA library by PCR (see T80195-96). These short
CC fragments were used to screen the cDNA library to obtain larger
CC fragments, and missing 5' ends were obtained by RACE. A
CC recombinant host cell, transfected with a vector comprising a cpk
CC nucleic acid is claimed. Novel phosphatidyl inositol 3-kinases
CC can be used to screen for agonists/antagonists of activity and
CC in a claimed method of treating a disorder caused by dysregulation
CC of a growth factor activation signalling cascade. Antagonists
CC may reduce Ras activation allowing treatment of proliferative
CC disorders such as atherosclerosis, inflammatory joint disease,
CC psoriasis, restenosis following angioplasty, and cancer.
SQ Sequence 6831 BP; 1829 A; 1685 C; 1669 G; 1648 T;

Query Match 58.1%; Score 25; DB 1; Length 6831;
Best Local Similarity 60.5%; Pred. No. 0.65;
Matches 23; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 6 YGAYYTRGNCARGAARTGCCRAATCDATRTGAA 43
DB 4544 TGGCATTCGCCAGAACTTCCAAAGTCATRTGAA 4507

RESULT 2
ID V74539
AC V74539; standard; DNA; 4488 BP.
DE 16-MAR-1999 (first entry)
KW Staphylococcus aureus contig SEQ ID #228.
KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.

PT	Isolated DNA encoding human SDF-5 protein - useful for controlling
PT	growth, differentiation etc. of cells, particularly of chondrocytes
PT	for treatment of arthritis etc., also pancreatic cells
PS	Claim 1; Page 35-57; 68pp; English.
CC	The sequence is that encoding human SDF-5, a member of the Frazzled
CC	protein family. Cells transformed with a vector containing the sequence
CC	are used to regulate genes, particularly pancreatic genes, or
CC	in combination with bone morphogenic protein 2 (BMP2), to
CC	increase differentiation of progenitor cells into chondrocytes.
CC	The protein may be used to treat osteoarthritis, rheumatoid
CC	arthritis, or articular cartilage defects, also to increase/inhibit cell
CC	formation, growth, differentiation, proliferation and/or maintenance in
CC	many other organs or tissues, e.g. for prevention or treatment of
CC	pancreatic cancer, diabetes (by inducing de novo formation of islet
CC	cells), other tissue defects, also to improve healing of wounds and to
CC	increase survival of nervous system cells, e.g. in cases of transplants
CC	The coding sequence can be used in gene therapy, and its fragments to
CC	detect related mRNA, while the protein is also used to generate
CC	antibodies, useful for affinity purification and as immunoassay
CC	reagents. Many other potential uses/activities for the gene and its
CC	encoded are contemplated but not exemplified, e.g. as cytokines,
CC	immunosuppressants or immunostimulants, regulators of haemopoiesis,
CC	as fertility-control agents, haemostatic or thrombolytic agents,
CC	anti-inflammatory agents, antimicrobials, modulators of biorthlyms
CC	and many more.
SQ	Sequence 2027 BP; 496 A; 583 C; 458 G; 490 T;
Query Match	48.4%; Score 20.8; DB 1; Length 2027;
Best Local Similarity	47.6%; Pred. No. 23;
Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;	
OY	1 GGNGAGCAATTCGCGCAGGARAAATGCCCAATCTCATRTGGA 42
Db	858 GGAGTAACCTACATCAATCAACGAGATACCAATAATCATCTTGG 899
V34257	: : : : : : :
RESULT 7	V34257 standard; DNA: 1960 BP.
ID	28-JAN-1999 (first entry)
AC	V34257:
DE	Human secreted protein gene 46 clone HLHD71.
KW	Human; secreted protein; fusion protein; gene therapy; protein therapy;
KW	diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
KW	developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
KW	immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
KW	inflammation; ischemic shock; Alzheimer's disease; restenosis; AIDS;
KW	cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
KW	osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
KW	endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
OS	Homo sapiens.
PV	WO9839446-A2.
PD	11-SEP-1998.
DD	06-MAR-1998; U04492.
PR	07-MAR-1997; US-038621.
PR	07-MAR-1997; US-040161.
PR	07-MAR-1997; US-040162.
PR	07-MAR-1997; US-040163.
PR	07-MAR-1997; US-040333.
PR	07-MAR-1997; US-040334.
PR	07-MAR-1997; US-040336.
PR	07-MAR-1997; US-040626.
PR	11-APR-1997; US-043311.
PR	11-APR-1997; US-043312.
PR	11-APR-1997; US-043313.
PR	11-APR-1997; US-043314.
PR	11-APR-1997; US-043315.
PR	11-APR-1997; US-043368.
PR	11-APR-1997; US-043369.
PR	11-APR-1997; US-043376.
PR	11-APR-1997; US-043578.
PR	11-APR-1997; US-043580.

PR 11-APR-1997; US-043669.
 PR 11-APR-1997; US-043670.
 PR 11-APR-1997; US-043671.
 PR 11-APR-1997; US-043672.
 PR 11-APR-1997; US-043673.
 PR 23-MAY-1997; US-047492.
 PR 23-MAY-1997; US-047500.
 PR 23-MAY-1997; US-047501.
 PR 23-MAY-1997; US-047502.
 PR 23-MAY-1997; US-047503.
 PR 23-MAY-1997; US-047581.
 PR 23-MAY-1997; US-047582.
 PR 23-MAY-1997; US-047583.
 PR 23-MAY-1997; US-047584.
 PR 23-MAY-1997; US-047585.
 PR 23-MAY-1997; US-047586.
 PR 23-MAY-1997; US-047587.
 PR 23-MAY-1997; US-047588.
 PR 23-MAY-1997; US-047589.
 PR 23-MAY-1997; US-047590.
 PR 23-MAY-1997; US-047592.
 PR 23-MAY-1997; US-047593.
 PR 23-MAY-1997; US-047594.
 PR 23-MAY-1997; US-047595.
 PR 23-MAY-1997; US-047596.
 PR 23-MAY-1997; US-047597.
 PR 23-MAY-1997; US-047598.
 PR 23-MAY-1997; US-047599.
 PR 23-MAY-1997; US-047600.
 PR 23-MAY-1997; US-047601.
 PR 23-MAY-1997; US-047612.
 PR 23-MAY-1997; US-047613.
 PR 23-MAY-1997; US-047614.
 PR 23-MAY-1997; US-047615.
 PR 23-MAY-1997; US-047617.
 PR 23-MAY-1997; US-047618.
 PR 23-MAY-1997; US-047632.
 PR 23-MAY-1997; US-047633.
 PR 06-JUN-1997; US-048964.
 PR 06-JUN-1997; US-048974.
 PR 22-AUG-1997; US-056630.
 PR 22-AUG-1997; US-056631.
 PR 22-AUG-1997; US-056632.
 PR 22-AUG-1997; US-056633.
 PR 22-AUG-1997; US-056636.
 PR 22-AUG-1997; US-056637.
 PR 22-AUG-1997; US-056662.
 PR 22-AUG-1997; US-056664.
 PR 22-AUG-1997; US-056845.
 PR 22-AUG-1997; US-056862.
 PR 22-AUG-1997; US-056864.
 PR 22-AUG-1997; US-056872.
 PR 22-AUG-1997; US-056874.
 PR 22-AUG-1997; US-056875.
 PR 22-AUG-1997; US-056876.
 PR 22-AUG-1997; US-056877.
 PR 22-AUG-1997; US-056878.
 PR 22-AUG-1997; US-056879.
 PR 22-AUG-1997; US-056880.
 PR 22-AUG-1997; US-056881.
 PR 22-AUG-1997; US-056882.
 PR 22-AUG-1997; US-056884.
 PR 22-AUG-1997; US-056886.
 PR 22-AUG-1997; US-056887.
 PR 22-AUG-1997; US-056888.
 PR 22-AUG-1997; US-056889.
 PR 22-AUG-1997; US-056892.
 PR 22-AUG-1997; US-056893.
 PR 22-AUG-1997; US-056894.
 PR 22-AUG-1997; US-056903.
 PR 22-AUG-1997; US-056908.
 PR 22-AUG-1997; US-056909.
 PR 22-AUG-1997; US-056910.
 PR 22-AUG-1997; US-056911.

PR 05-SEP-1997; US-057650.
 PR 05-SEP-1997; US-057651.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA,
 PI Feng P, Ferlie AM, Fischer CL, Graves KA, Greene JW, Hu JS,
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z,
 DR WPI; 98-60987/51.
 DR P-PSDB; W75160.
 PT New isolated human genes and the secreted polypeptides they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 PS Claim 1: Page 259-260; 447pp: English.
 CC This sequence represents a nucleic acid molecule which encodes a secreted
 CC human protein. The gene number, and the clone it is derived from, are
 CC detailed in the descriptor line. The gene can be used to generate fusion
 CC proteins by linking to the gene to a human immunoglobulin Fc portion
 CC (e.g. V34154) for increasing the stability of the fused protein as
 CC compared to the human protein only.
 CC The invention relates to 70 novel genes and their fragments (nucleic acid
 CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which
 CC are useful for preventing, treating or ameliorating medical conditions
 CC e.g. by protein or gene therapy. Also, pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the 70 polynucleotides, based on
 CC which tissues they are most highly expressed in (see V34154 for described
 CC uses).
 SQ Sequence 1960 BP; 504 A; 542 C; 436 G; 476 T;

Query Match 48.4%; Score 20.8; DB 1; Length 1960;
 Best Local Similarity 47.6%; Pred. No. 22;
 Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
 OY 1 GGGGAYGAYTTCGNCAGBARCGCRRARCCDATTGTA 42
 DB 767 GGAGATACCTACATACCGAGATACCAATCATCTGGA 808

RESULT 8
 V34199
 ID V34199 standard; DNA; 1869 BP.
 AC V34199;
 DT 28-JAN-1999 (first entry)
 DE Human secreted protein gene 46 clone HLHDY31.
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;
 KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 OS Homo sapiens.
 PN W09839446-K2.
 PD 11-SEP-1998.
 PE 06-MAR-1998; 004492.
 PR 07-MAR-1997; US-038621.
 PR 07-MAR-1997; US-040161.
 PR 07-MAR-1997; US-040162.
 PR 07-MAR-1997; US-040163.
 PR 07-MAR-1997; US-040333.
 PR 07-MAR-1997; US-040334.
 PR 07-MAR-1997; US-040336.
 PR 07-MAR-1997; US-040626.
 PR 11-APR-1997; US-043311.
 PR 11-APR-1997; US-043312.
 PR 11-APR-1997; US-043313.
 PR 11-APR-1997; US-043314.
 PR 11-APR-1997; US-043315.
 PR 11-APR-1997; US-043568.
 PR 11-APR-1997; US-043569.

DR 11-APR-1997; US-043576.
DR 11-APR-1997; US-043578.
PR 11-APR-1997; US-043580.
PR 11-APR-1997; US-043669.
PR 11-APR-1997; US-043670.
PR 11-APR-1997; US-043671.
PR 11-APR-1997; US-043672.
PR 11-APR-1997; US-043674.
PR 23-MAY-1997; US-047492.
PR 23-MAY-1997; US-047500.
PR 23-MAY-1997; US-047501.
PR 23-MAY-1997; US-047502.
PR 23-MAY-1997; US-047503.
PR 23-MAY-1997; US-047581.
PR 23-MAY-1997; US-047582.
PR 23-MAY-1997; US-047583.
PR 23-MAY-1997; US-047584.
PR 23-MAY-1997; US-047585.
PR 23-MAY-1997; US-047586.
PR 23-MAY-1997; US-047587.
PR 23-MAY-1997; US-047588.
PR 23-MAY-1997; US-047589.
PR 23-MAY-1997; US-047590.
PR 23-MAY-1997; US-047592.
PR 23-MAY-1997; US-047593.
PR 23-MAY-1997; US-047594.
PR 23-MAY-1997; US-047595.
PR 23-MAY-1997; US-047596.
PR 23-MAY-1997; US-047597.
PR 23-MAY-1997; US-047598.
PR 23-MAY-1997; US-047599.
PR 23-MAY-1997; US-047600.
PR 23-MAY-1997; US-047601.
PR 23-MAY-1997; US-047612.
PR 23-MAY-1997; US-047613.
PR 23-MAY-1997; US-047614.
PR 23-MAY-1997; US-047615.
PR 23-MAY-1997; US-047617.
PR 23-MAY-1997; US-047618.
PR 23-MAY-1997; US-047632.
PR 23-MAY-1997; US-047633.
PR 06-JUN-1997; US-048964.
PR 06-JUN-1997; US-048974.
PR 22-AUG-1997; US-056630.
PR 22-AUG-1997; US-056631.
PR 22-AUG-1997; US-056632.
PR 22-AUG-1997; US-056636.
PR 22-AUG-1997; US-056637.
PR 22-AUG-1997; US-056662.
PR 22-AUG-1997; US-056664.
PR 22-AUG-1997; US-056681.
PR 22-AUG-1997; US-056682.
PR 22-AUG-1997; US-056684.
PR 22-AUG-1997; US-056685.
PR 22-AUG-1997; US-056687.
PR 22-AUG-1997; US-056688.
PR 22-AUG-1997; US-056689.
PR 22-AUG-1997; US-056692.
PR 22-AUG-1997; US-056693.
PR 22-AUG-1997; US-056694.
PR 22-AUG-1997; US-056903.
PR 22-AUG-1997; US-056908.

[illegible]

CC murine SDF5. ATG-1622 polypeptides and polynucleotides are useful for
 CC diagnosing susceptibility to diseases by detecting mutations in the
 CC ATG-1622 gene and can diagnose diseases associated with ATG-1622
 CC imbalance. The polypeptides can be used to screen for agonists and
 CC antagonists which can be used in treatment to activate or inhibit
 CC ATG-1622 activity. The ATG-1622 polypeptide can be administered directly
 CC or as a vaccine to inoculate against disease. Diseases which can be
 CC diagnosed, prevented or treated by the ATG-1622 polypeptide or
 CC polynucleotides include heart disease, hypertension, kidney diseases,
 CC obesity, insulin resistance, lipodystrophy, diabetes and central nervous
 CC system (CNS) diseases. The ATG-1622 polypeptide is also useful for
 CC mapping the gene to a chromosome, allowing gene inheritance to be
 CC studied through linkage analysis.
 SQ Sequence 1984 BP; 477 A; 564 C; 452 G; 484 T;

Query Match 48.4%; Score 20.8; DB 1; Length 1984;
 Best Local Similarity 47.6%; Pred. No. 22;
 Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 1 GGNAGAYYTRCGNCARGAARATGCCRAARTCDATRTGSA 42
 DB 841 GGAGATACCTACATCAACGAGATACCAAAATCATCTCGA 882

RESULT 10
 X34662
 ID X34662 standard; cDNA; 1984 BP.

AC X34662;
 DT 01-JUL-1999 (first entry)
 DE Human secreted protein ligand ATG-1622 polypeptide encoding cDNA.
 KW Human; ATG-1622 polypeptide; secreted protein ligand; 7-TM receptor;
 KW flizzled family; heart disease; hypertension; kidney disease; obesity;
 KW insulin resistance; lipodystrophy; diabetes; CNS disease; ss.
 OS Homo sapiens.
 PN CA22929436-A.
 PD 21-NOV-1998.
 PF 23-APR-1998; 229436.
 PR 20-MAR-1998; US-044865.
 PR 21-MAY-1997; US-047251.
 PR 13-JUN-1997; US-874156.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Hu E. Zhu Y;
 DR WPI: 99-244679/21.
 DR P-PSDB: Y06923.

PT New secreted protein ligand polypeptide (ATG-1622) of 7-TM (flizzled
 PT family) receptors, useful for diagnosis, prevention and treatment of
 PS obesity, lipodystrophy, hypertension and heart disease
 PS Claim 4; Page 10; 42pp; English.
 CC This cDNA encodes a human ATG-1622 polypeptide, a secreted protein ligand
 CC for 7-TM (flizzled family) receptors. ATG-1622 polypeptides and
 CC polynucleotides are useful for diagnosing a disease or susceptibility to
 CC a disease by determining the presence/absence of a mutation in the ATG-
 CC 1622 gene, and/or analysing for the presence or amount of ATG-1622
 CC expressed in the patient. ATG-1622 expressing cells are useful for
 CC identifying modulators of the polypeptide which are can be used to treat
 CC conditions associated with a lack or excess of ATG-1622 polypeptide.
 CC Diseases diagnosed, prevented or treated include heart disease,
 CC hypertension, kidney diseases, obesity, insulin resistance,
 CC lipodystrophy, diabetes and CNS diseases. ATG-1622 protein may be useful
 CC for treating or preventing the onset of obesity. ATG-1622 polypeptides
 CC are also useful for mapping genes to chromosomes, allowing gene
 CC inheritance to be studied through linkage analysis.
 SQ Sequence 1984 BP; 477 A; 564 C; 452 G; 484 T;

Query Match 48.4%; Score 20.8; DB 1; Length 1984;
 Best Local Similarity 47.6%; Pred. No. 22;
 Matches 20; Conservative 11; Mismatches 11; Indels 0; Gaps 0;

OY 1 GGNAGAYYTRCGNCARGAARATGCCRAARTCDATRTGSA 42
 DB 841 GGAGATACCTACATCAACGAGATACCAAAATCATCTCGA 882

RESULT 11

ID 006088 standard; cDNA; 2289 BP.
 AC 006088;
 DT 31-JAN-1991 (first entry)
 DE Tumour cytotoxic factor II gene.
 KW TCF II; glycoprotein; fibroblasts; anticancer agent; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT cds 78..2249
 FT /tag= a

PN W09010651-A.
 PD 20-SEP-1990.
 PF 03-SEP-1990; J00314.
 PR 03-OCT-1989; JP-058631.
 PR 16-JAN-1990; JP-006692.
 PA (SNOW) SNOW BRAND MILK PRODUCTS.
 PI Higashio K; Mistuda S; Shima N; Itagaki Y; Nagao M.
 DR WPI: 90-304990/40.
 DR P-PSDB: R07144.

PT Human glycoprotein having antitumour activity - for inducing
 PT differentiation of leukemia cells and enhancing cell mediated
 PT immunity.
 PS Disclosure; Fig 15; 73pp; Japanese.
 CC The sequence was obtd. from a clone isolated from a cDNA library
 CC prepd. from RNA extd. from fibroblast IMR-90 cells (ATCC CCL-186).
 CC The DNA can be used to produce the TCF II by recombinant DNA
 CC technology. TCF II is an anticancer agent and induces
 CC differentiation of leukemia cells, enhances cell-mediated immunity,
 CC and accelerates the proliferation of human blood vessel endothelial
 CC cells and hepatic parenchymal cells.
 SQ Sequence 2289 BP; 703 A; 491 C; 523 G; 572 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2289;
 Best Local Similarity 52.5%; Pred. No. 34;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

OY 2 GGNAGAYYTRCGNCARGAARATGCCRAARTCDATRTGSA 41
 DB 1099 GCAGAGACCTACGAGAAATTTACTGCCGAATTCAGATCG 1138

RESULT 12

ID 010489 standard; DNA; 2187 BP.
 AC 010489;
 DT 16-APR-1991 (first entry)
 DE Hepatic parenchymal cell growth factor gene.
 KW Hepatic parenchymal cell growth factor; cirrhosis; ss.
 OS Homo sapiens.

FH Key Location/Qualifiers
 FT cds 1..2187
 FT /tag= a
 FT /product= hHGF
 FT /tag= b
 FT /product= hHGF active fragment
 FT /tag= c
 FT /product= hHGF active fragment
 FT /tag= c
 FT /product= hHGF active fragment

PN EP-412557-A.
 PD 13-FEB-1991.
 PF 10-AUG-1990; 115397.
 PR 11-AUG-1988; JP-209449.
 PR 03-APR-1990; JP-088592.
 PR 27-JUL-1990; JP-200898.
 PA (MIRU) MITSUBISHI KASEI CORP.
 PI Kitamura N; Miyazawa K; Daikuhara Y; Tsubouchi H; Naka D;
 PI Takahashi, Matsui R, Yoshiyama Y;

DR WPI: 91-045716/07.
 DR P-PSDB: R10656.
 PT Hepatic parenchymal cell growth factor - produced by recombinant
 PS methods, useful for e.g. treating cirrhosis
 CC Disclosure; fig 2; 46pp; English.
 CC This hepatic growth factor gene and its active fragments (see
 CC feature table) are contained within the expression plasmid,
 CC PKCHGF-2. This is used to transform host cells for recombinant
 CC expression of human hepatic parenchymal cell growth factor (hHGF).
 CC hHGF is useful in the treatment of hepatic diseases, e.g. cirr-
 CC hosis. It is produced in large amounts, stably using this method.
 SQ Sequence 2187 BP; 673 A; 456 C; 501 G; 557 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2187;
 Best Local Similarity 52.5%; Pred. No. 33;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 1037 GCAAGGACCTACGAGAAATTAATCTGCCGAATCCAGATGG 1076

RESULT 13
 ID Q12397 standard; CDNA; 2395 BP.
 AC Q12397;
 DT 17-SEP-1991 (first entry)
 DE Human hepatocyte growth factor HBC25 CDNA.
 KM Human hepatocyte growth factor; HGF; recombinant; HBC25; ss.
 PN J03130091-A.
 PD 03-JUN-1991.
 PF 05-JUN-1989; 142697.
 PR 05-JUN-1989; JP-142697.
 RA (TOYM) TOYORO KK.
 DR WPI: 91-204688/28.
 PT Recombinant human hepatocyte growth factor - used to produce
 PT polypeptide which promotes in-vitro growth of hepatocyte(s)
 PS Disclosure; fig 3(a); 20pp; Japanese.
 CC The complete sequence for the recombinant hHGF is presented in
 CC Q12399. The GF can be used in compns. for the treatment of
 CC hepatic disorders. See also Q12397-99.
 SQ Sequence 2395 BP; 748 A; 458 C; 501 G; 688 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2395;
 Best Local Similarity 52.5%; Pred. No. 34;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 465 GCAAGGACCTACGAGAAATTAATCTGCCGAATCCAGATGG 504

RESULT 14
 ID Q12398 standard; CDNA; 1661 BP.
 AC Q12398;
 DT 17-SEP-1991 (first entry)
 DE Human hepatocyte growth factor HAC19 CDNA.
 KM Human hepatocyte growth factor; HGF; recombinant; HAC19; ss.
 PN J03130091-A.
 PD 03-JUN-1991.
 PF 05-JUN-1989; 142697.
 PR 05-JUN-1989; JP-142697.
 RA (TOYM) TOYORO KK.
 DR WPI: 91-204688/28.
 PT Recombinant human hepatocyte growth factor - used to produce
 PT polypeptide which promotes in-vitro growth of hepatocyte(s)
 PS Disclosure; fig 3(b); 20pp; Japanese.
 CC The complete sequence for the recombinant hHGF is presented in
 CC Q12399. The GF can be used in compns. for the treatment of
 CC hepatic disorders. See also Q12397-99.

SQ Sequence 1661 BP; 526 A; 377 C; 363 G; 395 T;

Query Match 47.4%; Score 20.4; DB 1; Length 1661;
 Best Local Similarity 52.5%; Pred. No. 31;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 1064 GCAAGGACCTACGAGAAATTAATCTGCCGAATCCAGATGG 1103

RESULT 15
 ID Q12399 standard; DNA; 2187 BP.
 AC Q12399;
 DT 17-SEP-1991 (first entry)
 DE Human hepatocyte growth factor.
 KM Human hepatocyte growth factor; HGF; recombinant; ss.
 PN J03130091-A.
 PD 03-JUN-1991.
 PF 05-JUN-1989; 142697.
 PR 05-JUN-1989; JP-142697.
 RA (TOYM) TOYORO KK.
 DR WPI: 91-204688/28.
 DR P-PSDB: R12792.
 PT Recombinant human hepatocyte growth factor - used to produce
 PT polypeptide which promotes in-vitro growth of hepatocyte(s)
 PS Disclosure; fig 4(1-4); 20pp; Japanese.
 CC The GF can be used in compns. for the treatment of
 CC hepatic disorders. See also Q12397-99.
 SQ Sequence 2187 BP; 676 A; 460 C; 500 G; 551 T;

Query Match 47.4%; Score 20.4; DB 1; Length 2187;
 Best Local Similarity 52.5%; Pred. No. 33;
 Matches 21; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 GNGATGAYTTCGNCARGAARATGCCRAATCDAATRTGR 41
 DB 1037 GCAAGGACCTACGAGAAATTAATCTGCCGAATCCAGATGG 1076

Search completed: September 13, 1999, 16:22:08
 Job time: 4028 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 13, 1999, 16:14:01 ; Search time 849.94 Seconds

(without alignments)
99.794 Million cell updates/sec

Title: US-09-325-095-2729.SEQ

Sequence: 1 GCGAGATGATTCGNCARCA.....RTGCCRAATTCATRTGAA 43

Scoring table:

Search: 2546578 seqs, 98626752 residues

Database :

EST: *
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	22.8	53.0	289	44	AA728172	AA728172 Lmlv39p7/
2	22	51.2	384	27	AA028985	AA028985 2K11C01.r
3	22	51.2	599	38	AA801073	AA801073 EST190570
4	22	51.2	572	51	A1726663	A1726663 BNLG1578
5	21.8	50.7	458	23	H51655	H51655 yp81b08.r1
6	21.8	50.7	462	26	W77968	W77968 zd70e11.r1
7	21.8	50.7	641	48	A1559028	A1559028 AEMTBA58
8	21.6	50.2	474	33	AA391489	AA391489 LD10161.5
9	21.4	49.8	392	34	AA499435	AA499435 v185e01.r
10	21.4	49.8	378	34	AA512088	AA512088 v143n01.r
11	21.4	49.8	245	37	AA725990	AA725990 vU84h11.r
12	21.4	49.8	579	37	AA726611	AA726611 vU93d09.r
13	21.4	49.8	638	37	AA726685	AA726685 vU93h06.r
14	21.4	49.8	600	38	AA795990	AA795990 vs66d10.r
15	21.4	49.8	450	39	C88410	C88410 C88410 Carp
16	21.4	49.8	749	40	AA951519	AA951519 LD32028.5
17	21.4	49.8	551	40	AA951564	AA951564 LD32147.5
18	21.2	49.3	490	20	T41916	T41916 5179 Lambda
19	21.2	49.3	435	27	W99426	W99426 TgESTz201d1
20	21.2	49.3	460	35	C28257	C28257 C28257 Rice
21	21.2	49.3	460	35	C28712	C28712 C28712 Rice
22	21.2	49.3	433	39	AA856072	AA856072 w482e08.r
23	21.2	49.3	340	46	A1433117	A1433117 tH41f05.x
24	21.2	49.3	414	51	A1737481	A1737481 T110108e
25	20.8	48.4	446	20	M62196	M62196 EST00257 H1
26	20.8	48.4	406	23	H16121	H16121 Y128c05.r1
27	20.8	48.4	359	31	AA295018	AA295018 EST100474
28	20.8	48.4	302	32	AA330758	AA330758 EST34493
29	20.8	48.4	378	33	AA429960	AA429960 zW58g10.r
30	20.8	48.4	444	37	AA706119	AA706119 2159f03.s
31	20.8	48.4	515	42	A1105743	A1105743 SWAMCACC25
32	20.6	47.9	416	20	D24386	D24386 RICR1821A.R
33	20.6	47.9	429	20	D24456	D24456 RICR1950A.R
34	20.6	47.9	330	20	Z33790	Z33790 ATTS2881.Ye
35	20.6	47.9	430	23	T75952	T75952 10730 Lambd
36	20.6	47.9	544	24	N38187	N38187 19414 Lambd
37	20.6	47.9	334	35	C28727	C28727 C28727 Rice
38	20.6	47.9	462	36	C72767	C72767 C72767 Rice
39	20.6	47.9	287	40	C91804	C91804 C91804 Rice
40	20.6	47.9	327	50	A0064480	A0064480 A0064480
41	20.6	47.9	353	50	A0064483	A0064483 A0064483
42	20.6	47.9	260	51	A0068604	A0068604 A0068604
43	20.4	47.4	491	21	R13658	R13658 yf60f07.r1
44	20.4	47.4	448	22	R52797	R52797 yf99f12.r1
45	20.4	47.4	522	29	AA156676	AA156676 2118h04.S

ALIGNMENTS

RESULT 1
AA728172 289 bp mRNA EST 10-DEC-1998
LOCUS Lmlv39p7/352M Leishmania major RNA
DEFINITION Lmlv39p7/352M Leishmania major RNA
library from logarithmic stage (day 7) Leishmania major cDNA clone
352M 5', mRNA sequence.
ACCESSION AA728172
NID 93926020

VERSION AA728172.1 GI:3926020
 KEYWORDS EST.
 SOURCE Leishmania major.
 ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 REFERENCE 1 (bases 1 to 289)
 AUTHORS Almeida, R., Sampayo, I., Schneider, H. and Blackwell, J.M.
 TITLE Analysis of Leishmania major promastigote library from different stages of development
 JOURNAL Unpublished (1998)
 COMMENT On Jan 5, 1998 this sequence version replaced gi:2747129.

CONTACT: Blackwell JM
 Cambridge Institute for Medical Research
 Wellcome Trust/MRC Building, Addenbrooke's Hospital, Hills Road,
 Cambridge CB2 2XY, UK
 Tel: 01223 336 143
 Fax: 01223 331 206
 Email: jmb37@cus.cam.ac.uk
 PCR Primers
 FORWARD: GTAAACGACGCGCAGT
 BACKWARD: GGAACAGCATGACCATG
 Seq primer: AATTACCTCCTACTAAGG
 High quality sequence stop: 289.
 Location/Qualifiers
 1.289
 /organism="Leishmania major"
 /strain="LV39"
 /db_xref="taxon:5664"
 /clone="352W"
 /clone_1lb="Leishmania major promastigote full length cDNA library from logarithmic stage (day 7)"
 /cell_type="Promastigote"
 /note="Vector: Lambda Zap II; Site_1: XhoI; Site_2: NotI"

BASE COUNT 62 a 85 c 85 g 57 t
 ORIGIN

Query Match 53.0%; Score 22.8; DB 44; Length 289;
 Best Local Similarity 52.4%; Pred. No. 46;
 Matches 22; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTRCGNCGARARARATGCCAARTCCATRTGGA 42
 DB 55 GGTCAGACCTTCGCCAGAGAAATACGGAATACGCCGTGGA 96

RESULT 2
 LOCUS AA028985 384 bp mRNA EST 19-AUG-1996
 DEFINITION ZK11C01.1 Soares_pregnant_uterus_NbHPU Homo sapiens cDNA clone IMAGE:470208 5' similar to gb:x74929 KERATIN, TYPE II CYTOSKELETAL 8 (HUMAN)), mRNA sequence.
 ACCESSION AA028985
 NID 91496387
 VERSION AA028985.1 GI:1496387
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 384)
 AUTHORS Hillier, L., Clark, N., Dubugue, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisan, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.wustl.edu
 This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.lit.gov) for further information.
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 349.
 Location/Qualifiers
 1..384
 /organism="Homo sapiens"
 /db_xref="GDB:3756210"
 /db_xref="taxon:9606"
 /clone="IMAGE:470208"
 /clone_1lb="Soares_pregnant_uterus_NbHPU"
 /sex="female"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Organ: uterus; Vector: pT73-Pac; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', AACTGGAAGATTCGCGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by M. Fatima Bonaldo."

BASE COUNT 105 a 94 c 119 g 66 t
 ORIGIN

Query Match 51.2%; Score 22; DB 27; Length 384;
 Best Local Similarity 47.6%; Pred. No. 97;
 Matches 20; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

OY 1 GGNGAYGAYTRCGNCGARARARATGCCAARTCCATRTGGA 42
 DB 289 GGGAATGACCTCGCGGACAAAGACCTGAGATCTGTGAA 350

RESULT 3
 LOCUS AA801073 599 bp mRNA EST 30-APR-1998
 DEFINITION EST190570 Normalized rat ovary, Bento Soares Rattus sp. cDNA clone ROVA043 5' end, mRNA sequence.
 ACCESSION AA801073
 NID 92864028
 VERSION AA801073.1 GI:2864028
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 REFERENCE 1 (bases 1 to 599)
 AUTHORS Lee, N.H., Gloder, A., Chandra, I., Mason, T.M., Quackenbush, J., Kerlavage, A.R. and Adams, M.D.
 TITLE Rat Genome Project: Generation of a Rat EST (REEST) Catalog & Rat Gene Index
 JOURNAL Unpublished (1998)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1398005.

CONTACT: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712, Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13 Reverse.
 Location/Qualifiers
 1..599
 /organism="Rattus sp."
 /db_xref="ATCC (lnhost):2008443"

```

BASE COUNT      154 a      162 c      157 g      126 t
ORIGIN
      /db_xref="taxon:10118"
      /map="12g"
      /clone="KOVAA34"
      /clone_1b="Normalized rat ovary, Bento Soares"
      /note="Organ: ovary; Vector: pT7mpac; Site_1: ECORI
      Site_2: NotI"

```

Query Match	51.2%	Score 22	DB 38	Length 599
Best Local Similarity	47.6%	Pred No	1e+02	
Matches	20	Conservative	12	Mismatches 10; Indels 0; Gaps 0;

OY 1 GGAGGAYTTRCGCARGAARAARTGCCRAATCDATRTGGA 42
 ||-||:-||-||-||-||-||-||-||-||-||-||-||-||-||
Db 359 GGAGTGACTTACCTCAAGAGAACAACGNAAGGAATGAA 400

RESULT	4
A1726663/c	
LOCUS	A1726663 572 bp mRNA EST 11-JUN-1999
DEFINITION	BNIHG15782 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to ADENYLOSUCCINATE SYNTHETASE PRECURSOR (IMP--ASPARTATE LIGASE) g11161657 (U49389) adenylosuccinate synthetase [Arabidopsis thaliana], mRNA sequence.
ACCESSION	A1726663
NID	G5045515
VERSION	A1726663.1 GI:5045515
KEYWORDS	EST.
SOURCE	upland cotton.
ORGANISM	Gossypium hirsutum

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 572)	Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.	ESTs from developing cotton fiber	Unpublished (1999)	On May 18, 1998 this sequence version replaced g1:3138270.

Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burrb@bnl.bnl.gov
Seq primer: T3 primer.
Location/Qualifiers
1. 572

BASE COUNT	128 a	117 c	142 g	182 t	3 others
ORIGIN	/note="Vector: pBluescript II KS+"				

Query Match	51.2%	Score 22	DB 51	Length 572
Best Local Similarity	57.1%	Pred No. 1e+02		
Matches 20	Conservative 9	Mismatches 6	Indels 0	Gaps 0

RESULT	5		
H51655/c			
LOCUS	H51655	458 bp	mRNA
DEFINITION	y881808.r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone	EST	18-SEP-1995
IMAGE:193815	5'	similar to contains Alu repetitive element',	mRNA
sequence.			
unlabeled			
accession			

ACCESSION	H51655
NID	g991496
VERSION	H51655.1
KEYWORDS	GI:991496
SOURCE	EST.
ORGANISM	human.
	Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS Hillier, L., Clark, N., Dubouque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucabza, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.
TITLE The WashU-Merck EST Project
JOURNAL unpublished (1995)
COMMENT On Sep 21, 1992 this sequence version replaced gi:276010.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estewatson.wustl.edu
 Insert Size: 708
 High quality sequence stops: 314
 Source: IMAGE Consortium, LINTL
 This clone is available royalty-free through LINTL ; contact the
 IMAGE Consortium (info@image.lintl.gov) for further information.
 Insert Length: 708 Std Error: 0.00
 Seq Primer: M13Rpi
 High quality sequence stop: 334.
 Location/Qualifiers
 1..458

```

/organism="Homo sapiens"
/db_xref="GDB:3762836"
/db_xref="taxon:9606"
/clone="IMAGE:193815"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker. Site_1: Pac I; Site_2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dC) primer [5' AAGTCGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'] , double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

```

Query Match	50.7%	Score 21.8;	DB 23;	Length 458;
Best Local Similarity	55.3%	Pred. No. 1.2e+02;		
Matches 21; Conservative	9;	Mismatches 8;	Indels 0;	Gaps 0;

QY 5 AAGCAATTCGACGATGAGAAATGCGCCCAATATATATGTGA 42
 145 AAGATTTGCAAGCATGAAATGGGCCCAATAATATATATGTGA 108

RESULT	6
W77968	
LOCUS	W77968 462 bp mRNA EST 17-OCT-1996

```

DEFINITION      IMGE:J46028.5', mRNA sequence.
ACCESSION       w77968
NID             g1388502
VERSION         w77968.1
KEYWORDS        GI:1388502
SOURCE          #human.
ORGANISM        Homo sapiens
REFERENCE       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS         Eutheria; Primates; Catarrhini; Homnidae; Homo.
TITLE           1 (bases 1 to 462)
JOURNAL         Hillier,L., Clark,N., Dubaque,T., Elliston,R., Hawkins,M.,
COMMENT         Holman,M., Hultman,M., Kucaba,T., Le,M., Lenono,G., Marra,M.,
                Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
                Trevasis,E., Waterston,R., Williamson,A., Wohldmann,P. and
                Wilson,R.
                The Wash-Merck EST Project
                Unpublished (1995)
                On Apr 14, 1993 this sequence version replaced gi:838110.

FEATURES
    source
        Location/Qualifiers
            1..462
                /organism="Homo sapiens"
                /db_xref="GDB:1271403"
                /db_xref="taxon:9606"
                /clone="IMAGE:346028"
                /clone_lib="Soares_fetal_heart_NbHH19W"
                /sex="unknown"
                /dev_stage="19 weeks"
                /lab_host="DH10B (ampicillin resistant)"
                /note="Organ: heart; Vector: pUT3D (pharmacia) with a
                modified polylinker; Site.1: Not I; site.2: Eco RI; 1st
                strand cDNA was primed with a Not I - oligo(dT) primer [5'
                TGTTACCATCTGTGAAGTGCGAGCCGCCCATCTTTTTTTTTTTT 3'] ,
                double-stranded cDNA was size selected, ligated to Eco RI
                adapters (pharmacia), digested with Not I and cloned into
                the Not I and Eco RI sites of a modified pUT3 vector
                (Pharmacia). Library went through one round of
                normalization to a Cot = 5. Library constructed by
                M.Felipe Bonaldo. This library was constructed from the
                same fetus as the fetal lung library, Soares fetal lung
                NbHL19W."
BASE COUNT      128 a 133 c 106 g 91 t 4 others
ORIGIN          NBHL19W.
Query Match     50.7%; Score 21.8; DB 26; Length 462;
Best Local Similarity 48.8%; Pred.No.1.2e+02;
Matches 21; Conservative 11; Mismatches 11; Indels 0; Gaps 0;
QY              1 GGNGAYGAYTTCGCAGCARGAARATGCCTCAATCTGARTGRAA 43
Db              1 |||::|||::|||::|||::|||::|||::|||::|||::|||
                  318 GGAGATAACTTCATCAACGAGATTACCAAAATTCCTCGGAA 360

RESULT 7
LOCUS      A1559028               641 bp      mRNA
DEFINITION AEMTBAS5 Aedes aegypti MT psport Library Aedes aegypti cDNA clone
BAES 5',   mRNA sequence.
ACCESSION  A1559028
NID        94511742

```

```

VERSION      AI559028.1  GI:4511742
KEYWORDS
SOURCE       yellow fever mosquito.
ORGANISM     Aedes aegypti
              Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
              Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
              Culicidae; Culicidae; Aedes.
REFERENCE    1 (bases 1 to 641)
AUTHORS      Gill,S.S., Ross,L.S. and Wadlak,H.
TITLE        Expressed sequence tags of cDNA clones from an enriched Malpighian
              tubule and gut library from Aedes aegypti
              unpublished (1999)
JOURNAL      On Mar 24, 1999 this sequence version replaced.
COMMENT

FEATURES
    source
        1..641
        /organism="Aedes aegypti"
        /db_xref="taxon:7159"
        /clone="BA58"
        /clone_id="Aedes aegypti MT pSPORT Library"
        /sex="female"
        /tissue_type="malpighian tubules and gut"
        /dev_stage="adult"
        /lab_host="DH10B"
        /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
        cDNA was cloned into the SalI/NotI sites of pSPORT1."

BASE COUNT   150 a      153 c      175 g      163 t

ORIGIN
Query Match      50.7%; Score 21.8; DB 48; Length 641;
Best Local Similarity 51.2%; Pred.No. 1.2e+02;
Matches 21; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Oy      2  GNGAGATGTTTCGNCARARARATGCCAAATCATRIGRA 42
          |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      468 GAGATGCTGCGCCAGTAAAGTGTCCAACTGATGCTGA 508

RESULT      8
LOCUS       AA391489              474 bp      mRNA      EST      28-NOV-1998
DEFINITION LD10161:prIME LD Drosophila melanogaster embryo Bluescript
ACCESSION   AA391489
            AA391489
            92044464
            92044464
VERSION     AA391489.1  GI:2044464
KEYWORDS    EST
SOURCE      fruit fly
ORGANISM    Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
            1 (bases 1 to 474)
            Harvey,D., Hong,L., Evans,Holm,M., Pendleton,J., Su,C.,
            Brokstein,P., Lewis,S. and Rubin,G.M.
            BGGP/HMT Drosophila EST Project
            unpublished (1997)
JOURNAL     On Sep 12, 1996 this sequence version replaced gi:1397644.
COMMENT
TITLE       Contact: Harvey, D
            G.M. Rubin-Molecular and Cell Biology
            University of California Berkeley
            539 USA, Berkeley, CA 94720-3200, USA
JOURNAL
COMMENT

```

Fax: 510 643 9947

Email: <http://www.fruitfly.org/EST>, est@fruitfly.berkeley.edu

hit genomic sequence AC001661

Plate: 101 row: F column: 1

High quality sequence stop: 419.

FEATURES

Source

Location/Qualifiers
1..474
/organism="Drosophila melanogaster"
/db_xref="BDGP_EST:BDc1009432"
/db_xref="taxon:7227"
/clone="1D10161"
/clone_lib="1D Drosophila melanogaster embryo Bluescript"
/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: Bluescript SK; Site: 1;
EcoRI; Site 2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in Bluescript SK(+/-)"

BASE COUNT

173 a 93 c 88 g 120 t

ORIGIN

Query Match 50.2%; Score 21.6; DB 33; Length 474;
Best Local Similarity 51.4%; Pred. No. 1.4e+02;
Matches 19; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

Qy 4 GAGGAYTCGNCARGARARATGCCAATCCATRTG 40

Db 408 GCTGATTACAGCAGAGAGATTGGAATCTTATG 444

RESULT 9

AA49435 392 bp mRNA EST 01-JUL-1997
LOCUS v185e01.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:919032 5' similar to gb:X07696 KERATIN, TYPE I CYTOSKELETAL
15 (HUMAN); mRNA sequence.

ACCESSION

AA49435 92234402

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 392)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1404967.

FEATURES

Source

Location/Qualifiers
1..392
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"

/map="xg28"
/clone="IMAGE:919032"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pluscript SK; Site: 1;
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTGTGTTTGT 3'

BASE COUNT

113 a 105 c 103 g 70 t

ORIGIN

Query Match 49.8%; Score 21.4; DB 34; Length 392;
Best Local Similarity 46.5%; Pred. No. 1.6e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

Qy 1 GAGGAYTCGNCARGARARATGCCAATCCATRTG 43

Db 207 GGAGTGACTTCTCAAGAACACGAGAGAAATGAAA 249

RESULT 10

AA512088 378 bp mRNA EST 08-JUL-1997
LOCUS v143h01.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:931825 5' similar to gb:X07696 KERATIN, TYPE I CYTOSKELETAL
15 (HUMAN); gb:L00193 Mouse epidermal keratin type I intermediate
filament (MOUSE); mRNA sequence.

ACCESSION

AA512088 92249942

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 378)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1400805.

FEATURES

Source

Location/Qualifiers
1..378
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:931825"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"

/lab.host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site: 1;
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."
 BASE COUNT 110 a 99 c 102 g 67 t
 ORIGIN

Query Match 49.8%; Score 21.4; DB 34; Length 378;
 Best Local Similarity 46.5%; Pred. No. 1.5e+02;
 Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNAGAYTTCRCNCARGAARATCCRAARTCDATRTGAA 43
 DB 192 GGAGATGACTTATCTCAAGAGAACCCAGAGGAATGAAA 234

RESULT 11
 LOCUS AA725990 245 bp mRNA EST 02-JAN-1998
 DEFINITION vu84h11.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1198149 5' similar to gb:J00124.cd61 KERATIN, TYPE I
 CYTOSKELETAL 14 (HUMAN); gb:M13806 Mouse Keratin (MOUSE);, mRNA
 sequence.

ACCESSION AA725990
 MID 92743697
 VERSION AA725990.1 GI:2743697
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 245)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On May 18, 1995 this sequence version replaced gi:811118.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:645245
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 173.
 Location/Qualifiers

FEATURES
 source 1..245
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:1198149"
 /clone_lib="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor

BASE COUNT 69 a 66 c 74 g 36 t
 ORIGIN

Query Match 49.8%; Score 21.4; DB 37; Length 245;
 Best Local Similarity 46.5%; Pred. No. 1.5e+02;
 Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNAGAYTTCRCNCARGAARATCCRAARTCDATRTGAA 43
 DB 60 GGAGATGACTTATCTCAAGAGAACCCAGAGGAATGAAA 102

RESULT 12
 LOCUS AA726611 579 bp mRNA EST 02-JAN-1998
 DEFINITION vu93d09.r1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
 IMAGE:1209713 5' similar to gb:X07696 KERATIN, TYPE I CYTOSKELETAL
 15 (HUMAN);, mRNA sequence.

ACCESSION AA726611
 MID 92744318
 VERSION AA726611.1 GI:2744318
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 579)
 Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
 Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
 Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
 Waterston, R.

TITLE The WashU-HMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On May 5, 1995 this sequence version replaced gi:797825.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:646057
 Seq primer: -28m13 rev1 ET from Amersham
 High quality sequence stop: 440.
 Location/Qualifiers

FEATURES
 source 1..579
 /organism="Mus musculus"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /map="6"
 /clone="IMAGE:1209713"
 /clone_lib="Stratagene mouse skin (#937313)"
 /sex="females"
 /tissue_type="whole skin"
 /dev_stage="11 weeks old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: skin; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
 Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
 adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTCTTTTCTTTT 3'."

BASE COUNT 173 a 149 c 151 g 106 t
 ORIGIN

Query Match 49.8%; Score 21.4; DB 37; Length 579;

Best Local Similarity 46.5%; Pred. No. 1.7e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARGARARATCCCAATCDATRTGRGA 43
Db 393 GGAGATGACTTATCTCAAGAGAACCAAGAGAGAAATGAAA 435

RESULT 13

AA726685 638 bp mRNA EST 02-JAN-1998
LOCUS AA726685
DEFINITION v093h06.r1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone

*IMAGE:1209755.5' similar to gb:X07696 KERATIN, TYPE I CTOSKELETAL
15 (HUMAN); gb:M1805 Mouse type I epidermal keratin mRNA, clone
pMSC-50, 3' (MOUSE);, mRNA sequence.

ACCESSION AA726685.1 GI:2744392
NID 92744392
VERSION AA726685
KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS

Matta,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,U., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The Washu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On May 5, 1995 this sequence version replaced gi:798153.

FEATURES
source
Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:646099
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 445.
Location/Qualifiers
1..638
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1209755"
/clone_1lb="Stratiogene mouse skin (#937313)"
/sex="females"
/tissue-type="whole skin"
/dev-stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site:1;
EcoRI; Site:2; XhoI; Cloned unidirectionally. Primer:
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3''

BASE COUNT 175 a 180 c 176 g 107 t

ORIGIN

Query Match 49.8%; Score 21.4; DB 37; Length 638;
Best Local Similarity 46.5%; Pred. No. 1.8e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARGARARATCCCAATCDATRTGRGA 43
Db 218 GGAGATGACTTATCTCAAGAGAACCAAGAGAGAAATGAAA 260

RESULT 14
LOCUS AA795990/c
DEFINITION

AA795990 600 bp mRNA EST 09-FEB-1998
v566d10.r1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1151251.5' similar to gb:X07696 KERATIN, TYPE I CTOSKELETAL
15 (HUMAN);, mRNA sequence.

ACCESSION AA795990.1 GI:2858945
NID 92858945
VERSION AA795990
KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)

AUTHORS

Matta,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,U., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The Washu-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2044105.

FEATURES
source
Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:624459
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 364.
Location/Qualifiers
1..600
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1151251"
/clone_1lb="Stratiogene mouse skin (#937313)"
/sex="females"
/tissue-type="whole skin"
/dev-stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site:1;
EcoRI; Site:2; XhoI; Cloned unidirectionally. Primer:
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGGCGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3''

BASE COUNT 119 a 143 c 170 g 168 t

ORIGIN

Query Match 49.8%; Score 21.4; DB 38; Length 600;
Best Local Similarity 46.5%; Pred. No. 1.7e+02;
Matches 20; Conservative 12; Mismatches 11; Indels 0; Gaps 0;

QY 1 GGNGAYGAYTTCGNCARGARARATCCCAATCDATRTGRGA 43
Db 112 GGAGATGACTTATCTCAAGAGAACCAAGAGAGAAATGAAA 70

RESULT 15
LOCUS C88410
DEFINITION C88410 Carp albinase-induced 48 hr cDNA Cyp17a1 cDNA clone
M131, mRNA sequence.

```

ACCESSION      C88410
NID             92924692
VERSION        C88410.1  GI:2924692
KEYWORDS
SOURCE
ORGANISM        Cyprinus carpio
                common carp.
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
                Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
                Cyprinidae; Cyprininae; Cyprini.
REFERENCE
AUTHORS         Fujiki, R., Nakao, M., Shin, D. and Yano, T.
TITLE           1 (bases 1 to 450)
                Common carp genes differentially expressed in sodium
                alginate-elicited peritoneal exudate cells are identified by
                suppression subtractive hybridization technique
                Unpublished (1998)
JOURNAL
COMMENT         On Apr 14, 1993 this sequence version replaced gi:639236.

Contact: Kazuhiro Fujiki
Laboratory of Marine Biochemistry, Faculty of Agriculture
Kyushu University
Hakozaki 6-10-1, Higashi-ku, Fukuoka City, Fukuoka 812-8581, Japan
Tel: 81-92-642-2896
Fax: 81-92-642-2894
Email: fujiki@agr.kyushu-u.ac.jp.

FEATURES
  source
    1..450
    /organism="Cyprinus carpio"
    /db_xref="taxon:7962"
    /clone_id="Carp alginate-induced 48 hr cdna"
    /cell_type="peritoneal exudate cells"

BASE COUNT      104 a 117 c 119 g 110 t

ORIGIN
Query Match      49.8%; Score 21.4; DB 39; Length 450;
Best Local Similarity 48.8%; Pred. No. 1.7e+02;
Matches 20; Conservative 11; Mismatches 10; Indels 0; Gaps 0;

QY 2-GNAGAYGAYTTCGNCAGARARARTGCCRAARTCDARTGGA 42
Db 140 GACAGATTTCGCTCCACGACATGCCAAGTGAGTGTGGA 180

```

Search completed: September 13, 1999, 16:14:04
 Job time: 5463 sec